



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : G06F 19/00		A2	(11) International Publication Number: WO 00/42561
			(43) International Publication Date: 20 July 2000 (20.07.00)
(21) International Application Number: PCT/US00/01203			(72) Inventors; and
(22) International Filing Date: 18 January 2000 (18.01.00)			(75) Inventors/Applicants (for US only): CRAMERI, Andreas [CH/US]; Gehrenstrasse 3, CH-4153 Reinach (CH). STEMMER, Willem, P., C. [NL/US]; 108 Kathy Court, Los Gatos, CA 95030 (US). MINSHULL, Jeremy [GB/US]; 11 Homer Lane, Menlo Park, CA 94025 (US). BASS, Steven, H. [US/US]; 950 Parrot Drive, Hillsborough, CA 94010 (US). WELCH, Mark [US/US]; 25 Montalban Drive, Fremont, CA 94536 (US). NESS, Jon, E. [US/US]; 1220 N. Fair Oaks Avenue #3115, Sunnyvale, CA 94089 (US). GUSTAFSSON, Claes [SE/US]; 1813 Bayview Avenue, Belmont, CA 94002 (US). PATTEN, Phillip, A. [US/US]; Apartment 506, 2680 Fayette Drive, Mountain View, CA 94040 (US).
(30) Priority Data: 60/116,447 19 January 1999 (19.01.99) US 60/118,813 5 February 1999 (05.02.99) US 60/118,854 5 February 1999 (05.02.99) US 60/141,049 24 June 1999 (24.06.99) US 09/408,392 28 September 1999 (28.09.99) US 09/408,393 28 September 1999 (28.09.99) US 09/416,375 12 October 1999 (12.10.99) US 09/416,837 12 October 1999 (12.10.99) US			(74) Agents: QUINE, Jonathan, Alan; The Law Offices of Jonathan Alan Quine, P.O. Box 458, Alameda, CA 94501 (US) et al.
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 09/408,392 (CIP) Filed on 28 September 1999 (28.09.99)			(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(71) Applicant (for all designated States except US): MAXYGEN, INC. [US/US]; 515 Galveston Drive, Redwood City, CA 94063 (US).			Published Without international search report and to be republished upon receipt of that report.
(54) Title: OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION			
(57) Abstract			
<p>Methods of recombinating nucleic acids, including homologous nucleic acids, are provided. Families of gene shuffling oligonucleotides and their use in recombination procedures, as well as polymerase and ligase mediated recombination methods are also provided.</p>			
<p>Gene Family</p> <p>Synthesize Oligos</p> <p>Introduce into plant cells</p> <p>Grow up Plants</p> <p>Library of Genetically Enhanced Plants</p>			

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

FOR THE PURPOSES OF INFORMATION ONLY

OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION**5 CROSS REFERENCE TO RELATED APPLICATIONS**

This application is a continuation-in-part of "OLIGONUCLEOTIDE
10 MEDIATED NUCLEIC ACID RECOMBINATION" by Cramer et al., USSN 09/408,392,
filed September 28, 1999, which is a non-provisional of "OLIGONUCLEOTIDE
15 MEDIATED NUCLEIC ACID RECOMBINATION" by Cramer et al., USSN 60/118,813,
filed February 5, 1999 and which is also a non-provisional of "OLIGONUCLEOTIDE
20 MEDIATED NUCLEIC ACID RECOMBINATION" by Cramer et al., USSN 60/141,049,
filed June 24, 1999.

This application is also a continuation-in-part of "METHODS FOR MAKING
CHARACTER STRINGS, POLYNUCLEOTIDES AND POLYPEPTIDES HAVING
15 DESIRED CHARACTERISTICS" by Selifonov et al., attorney docket number 02-289-3US,
filed herewith, which is a continuation-in-part of "METHODS FOR MAKING
CHARACTER STRINGS, POLYNUCLEOTIDES AND POLYPEPTIDES HAVING
DESIRER CHARACTERISTICS" by Selifonov et al., USSN 09/416,375, filed October 12,
1999, which is a non-provisional of "METHODS FOR MAKING CHARACTER STRINGS,
20 POLYNUCLEOTIDES AND POLYPEPTIDES HAVING DESIRER
CHARACTERISTICS" by Selifonov and Stemmer, USSN 60/116,447, filed January 19,
1999 and which is also a non-provisional of "METHODS FOR MAKING CHARACTER
STRINGS, POLYNUCLEOTIDES AND POLYPEPTIDES HAVING DESIRER
CHARACTERISTICS" by Selifonov and Stemmer, USSN 60/118,854, filed February 5,
25 1999.

This application is also a continuation-in-part of co-filed application
"METHODS OF POPULATING DATA STRUCTURES FOR USE IN EVOLUTIONARY
SIMULATIONS" by Selifonov and Stemmer, Attorney Docket Number 3271.002WO0 (filed
by Majestic, Parsons, Siebert & Hsue) which is a continuation-in-part of "METHODS OF
30 POPULATING DATA STRUCTURES FOR USE IN EVOLUTIONARY SIMULATIONS"
by Selifonov and Stemmer, USSN 09/416,837, filed October 12, 1999.

This application is also related to "USE OF CODON VARIED
OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch et al.,
USSN 09/408,393, filed September 28, 1999.

5 (1996), "Construction And Evolution Of Antibody-Phage Libraries By DNA Shuffling",
 the enhancement of bioprocessing through recursive shuffling techniques, Crampet et al.,
 METABOLIC ENGINEERING provides, e.g., for the evolution of metabolic pathways and
 U.S. Pat. No. 5,837,458 METHODS AND COMPOSITIONS FOR CELLULAR AND
 inventors and their co-workers. In addition to the publications noted above, Minshall et al.,
 30 Applications of DNA shuffling technology have also been developed by the
 inventors of displayed peptides and antibodies.
 by repeated cycles of mutagenesis, shuffling and selection, as well as methods of generating
 e.g., in vitro and in vivo nucleic acid, DNA and protein shuffling in a variety of formats, e.g.,
 25 CHARACTERISTICS BY INTERATIVE SELECTION AND RECOMBINATION describe,
 METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESIRED
 FRAGMENTATION AND REASSEMBLY; Stemmer et al., U.S. Pat. No. 5,811,238
 Stemmer et al., U.S. Pat. No. 5,830,721 DNA MUTAGENESIS BY RANDOM
 Stemmer U.S. Patent No. 5,603,793 METHODS FOR IN VITRO RECOMBINATION;
 20 in vitro Recombination for Molecular Evolution, Proc. Natl. Acad. USA 91:10747-10751;
 370:389-391; Stemmer (1994), "DNA Shuffling by Random Fragmentation and Reassembly;
 shuffling. For example, Stemmer et al. (1994), "Rapid Evolution of a Protein" Nature
 A number of publications by the inventors and their co-workers describe DNA
 and their co-workers.
 15 apparatus for practicing these methods represent a pioneering body of work by the inventors
 therapeutic genes and encoded proteins. These methods, and related compositions and
 fast artificial evolution methodologies for generating improved industrial, agricultural, and
 generation, manipulation and selection. The inventors and their co-workers have developed
 DNA shuffling has provided a paradigm shift in recombinant nucleic acid
 10 BACKGROUND OF THE INVENTION
 otherwise reserves all copyright rights whatsoever.
 disclosure, as it appears in the Patent and Trademark Office patent file or records, but
 objection to the facsimile reproduction by anyone of the patent document or patent
 contains material which is subject to copyright protection. The copyright owner has no
 Pursuant to 37 C.F.R. 1.71(e), Applicants note that a portion of this disclosure
 5 COPYRIGHT NOTIFICATION
 §120, as appropriate.
 applications listed in this section, as provided for under 35 U.S.C. §119(e) and/or 35 U.S.C.
 The present application claims priority to and benefit of each of the

Nature Medicine 2(1):100-103 describe, e.g., antibody shuffling for antibody phage libraries. Additional details regarding DNA Shuffling can be found in WO95/22625, WO97/ 20078, WO96/33207, WO97/33957, WO98/27230, WO97/35966, WO98/ 31837, WO98/13487, WO98/13485 and WO989/42832, as well as a number of other publications by the inventors 5 and their co-workers.

A number of the publications of the inventors and their co-workers, as well as other investigators in the art also describe techniques which facilitate DNA shuffling, e.g., by providing for reassembly of genes from small fragments, or even oligonucleotides. For example, in addition to the publications noted above, Stemmer et al. (1998) U.S. Pat. No.

10 5,834,252 END COMPLEMENTARY POLYMERASE REACTION describe processes for amplifying and detecting a target sequence (e.g., in a mixture of nucleic acids), as well as for assembling large polynucleotides from nucleic acid fragments.

Review of the foregoing publications reveals that forced evolution by gene shuffling is an important new technique with many practical and powerful applications.

15 Thus, new techniques which facilitate gene shuffling are highly desirable. The present invention provides significant new gene shuffling protocols, as well as many other features which will be apparent upon complete review of this disclosure.

SUMMARY OF THE INVENTION

The invention provides oligonucleotide assisted shuffling of nucleic acids. 20 These oligonucleotide assisted approaches particularly facilitate family shuffling procedures, providing substantially simplified shuffling protocols which can be used to produce family shuffled nucleic acids without isolating or cloning full-length homologous nucleic acids. Furthermore, the oligonucleotide assisted approaches herein can even be extended to shuffling non-homologous nucleic acids, thereby accessing greater sequence space in 25 resulting recombinant molecules and, thus, greater molecular diversity. The techniques can also be combined with classical DNA shuffling protocols, such as DNase-mediated methods, or with other diversity generation procedures such as classical mutagenesis, to increase the versatility and throughput of these methods.

Several methods which are applicable to family shuffling procedures are 30 provided. In one aspect of these methods, sets of overlapping family gene shuffling oligonucleotides are hybridized and elongated, providing a population of recombined nucleic acids, which can be selected for a desired trait or property. Typically, the set of overlapping family shuffling gene oligonucleotides include a plurality of oligonucleotide member types which have consensus region subsequences derived from a plurality of homologous target

Typically, family gene shuffling oligonucleotides are provided by aligning homologous nucleic acid sequences to select conserved regions of sequence identity and regions of sequence diversity. A plurality of family gene shuffling oligonucleotides are synthesized (serially or in parallel) which correspond to at least one region of sequence diversity. In contrast, sets of fragments are provided by cleaving one or more homologous nucleic acids (e.g., with a DNase), or by synthesizing a set of oligonucleotides corresponding to a plurality of regions of at least one nucleic acid (typically oligonucleotides corresponding to a full-length nucleic acid provided as members of a set of nucleic acid fragments). In the shuffling procedure herein, these cleavage fragments can be used in conjunction with family gene shuffling oligonucleotides, e.g., in one or more recombination reaction. Recombinative methods of oligonucleotide shuffling are provided. As noted herein, recombinant nucleic acids generated synthetically using oligonucleotides can be cleaved and shuffled by standard nucleic acid shuffling methodologies, or the nucleic acids can be sequenced and used to design a second set of family shuffling oligonucleotides which are used to recombine the recombinant nucleic acids. Either, or both, of these recombinative techniques can be used for subsequent rounds of recombination and can also be used in conjunction with rounds of selection of recombinant products. Selection steps can follow one or several rounds of recombination, depending on the desired diversity of the recombinant

cross-over capability, codon-variation or selection, and the like. The population of recombinant nucleic acids can be denatured and reannealed, providing denatured recombinant nucleic acids which can then be reannealed. The resulting recombinant nucleic acids can also be selected. Any or all of these steps can be repeated relatively, providing for multiple recombination and selection events to produce a nucleic acid with a desired trait or property. In a related aspect, methods for introducing nucleic acid family diversity during nucleic acid recombination are performed by providing a composition having at least one set of fragmented nucleic acids which includes a population of family gene shuffling oligonucleotides and recombinant nucleic acids with at least one of the fragmented nucleic acids having a nucleic acid subsequence corresponding to the at least one family gene shuffling oligonucleotide. A recombinant nucleic acid having a nucleic acid shuffling is then regenerated, typically to encode a full-length molecule (e.g., a full-length protein).

nucleic acids (the more rounds of recombination which are performed, the more diverse the resulting population of recombinant nucleic acids).

The use of family gene shuffling oligonucleotides in recombination reactions herein provides for domain switching of domains of sequence identity or diversity between 5 homologous nucleic acids, e.g., where recombinants resulting from the recombination reaction provide recombinant nucleic acids with a sequence domain from a first nucleic acid embedded within a sequence corresponding to a second nucleic acid, e.g., where the region most similar to the embedded region from the second nucleic acid is not present in the recombinant nucleic acid.

10 One particular advantage of the present invention is the ability to recombine homologous nucleic acids with low sequence similarity, or even to recombine non-homologous nucleic acids. In these methods, one or more set of fragmented nucleic acids are recombined with a with a set of crossover family diversity oligonucleotides. Each of these crossover oligonucleotides have a plurality of sequence diversity domains corresponding to a 15 plurality of sequence diversity domains from homologous or non-homologous nucleic acids with low sequence similarity. The fragmented oligonucleotides, which are derived from one or more homologous or non-homologous nucleic acids can hybridize to one or more region of the crossover oligos, facilitating recombination.

Methods of family shuffling PCR amplicons using family diversity
20 oligonucleotide primers are also provided. In these methods, a plurality of non-homogeneous homologous template nucleic acids are provided. A plurality of PCR primers which hybridize to a plurality of the plurality of non-homogeneous homologous template nucleic acids are also provided. A plurality of PCR amplicons are produced by PCR amplification of the plurality of template nucleic acids with the plurality of PCR primers, which are then 25 recombined. Typically, sequences for the PCR primers are selected by aligning sequences for the plurality of non-homogeneous homologous template nucleic acids and selecting PCR primers which correspond to regions of sequence similarity.

A variety of compositions for practicing the above methods and which result from practicing the above methods are also provided. Compositions which include a library 30 of oligonucleotides having a plurality of oligonucleotide member types are one example. The library can include at least about 2, 3, 5, 10, 20, 30, 40, 50, 100 or more different oligonucleotide members. The oligonucleotide member types correspond to a plurality of subsequence regions of a plurality of members of a selected set of a plurality of homologous target sequences. The plurality of subsequence regions can include, e.g., a plurality of

two basic ways. First, a given sequence can be artificially recombined with another mutation and natural selection. Under artificial conditions, divergence occurs, e.g., in one of two or more descendant sequences diverge from a parent sequence over time, i.e., due to artificially, from a common ancestor sequence. During natural evolution, this occurs when nucleic acids are "homologous" when they are derived, naturally or

30

art.

Unless otherwise indicated, the following definitions supplement those in the

DEFINITIONS

Figure 3 is a schematic of a modular exon deletion/insertion library.

synthetic gene blending.

Figure 2 is a schematic of a low-homology shuffling procedure to provide for

using chimeraplasts.

Figure 1 is a schematic showing oligonucleotide-directed *in vivo* shuffling

BRIEF DESCRIPTION OF THE FIGURES

methods are also provided.

Invention are also provided. Uses for the compositions and kits herein for practicing the packaging materials, e.g., with instructional materials for practicing the methods of the invention are also provided.

Kits comprising the compositions of the invention, e.g., in containers, or other plurality of members of the selected set of homologous target sequences, and the like.

Synthesis reagents, buffers, salts, magnesium, one or more nucleic acid having one or more of the many other commercially available polymers) a recombinase, a nucleic acid of the many other commercially available polymers) a recombinase, a nucleic acid

Compositions optionally include components which facilitate recombination mediated recombination as noted herein are also a feature of the invention.

20

25

Libraries of nucleic acids and encoded proteins which result from practicing oligonucleotide each of the plurality of homologous oligonucleotides have at least one variant subsequence. In this embodiment, homologous region from the plurality of homologous target sequences. Member types can include a plurality of homologous oligonucleotides corresponding to a crossover oligonucleotides, domain switching oligonucleotides, etc.). The oligonucleotide included in the compositions of the invention (e.g., family shuffling oligonucleotides,

15

Any of the oligonucleotide types and sets described above, or elsewhere herein, can be least one subsequence from at least one of the selected set of homologous target sequences. The oligonucleotide member types typically each have a sequence identical to at least one sequence of the selected set of homologous target

10

sequence, as occurs, e.g., during typical cloning, to produce a descendent nucleic acid, or a given sequence can be chemically modified, or otherwise manipulated to modify the resulting molecule. Alternatively, a nucleic acid can be synthesized *de novo*, by synthesizing a nucleic acid which varies in sequence from a selected parental nucleic acid sequence. When there is 5 no explicit knowledge about the ancestry of two nucleic acids, homology is typically inferred by sequence comparison between two sequences. Where two nucleic acid sequences show sequence similarity over a significant portion of each of the nucleic acids, it is inferred that the two nucleic acids share a common ancestor. The precise level of sequence similarity which establishes homology varies in the art depending on a variety of factors. For purposes 10 of the present invention, cladistic intermediates (proposed sequences which share features of two or more related nucleic acids) are homologous nucleic acids.

For purposes of this disclosure, two nucleic acids are considered homologous where they share sufficient sequence identity to allow direct recombination to occur between the two nucleic acid molecules. Typically, nucleic acids utilize regions of close similarity 15 spaced roughly the same distance apart to permit recombination to occur. The recombination can be *in vitro* or *in vivo*.

It should be appreciated, however, that one advantage of certain features of the invention is the ability to recombine more distantly related nucleic acids than standard recombination techniques permit. In particular, sequences from two nucleic acids which are 20 distantly related, or even not detectably related can be recombined using cross-over oligonucleotides which have subsequences from two or more different non-homologous target nucleic acids, or two or more distantly related nucleic acids. However, where the two nucleic acids can only be indirectly recombined using oligonucleotide intermediates as set forth herein, they are considered to be "non-homologous" for purposes of this disclosure.

25 A "set" as used herein refers to a collection of at least two molecules types, and typically includes at least about, e.g., 5, 10, 50, 100, 500, 1,000 or more members, depending on the precise intended use of the set.

A set of "family gene shuffling oligonucleotides" is a set of synthesized 30 oligonucleotides derived from a selected set of homologous nucleic acids. The oligonucleotides are derived from a selected set of homologous nucleic acids when they (individually or collectively) have regions of sequence identity (and, optionally, regions of sequence diversity) with more than one of the homologous nucleic acids. Collectively, the oligonucleotides typically correspond to a substantial portion of the full length of the homologous nucleic acids of the set of homologous nucleic acids, e.g., the oligonucleotides

detected by cleaving one or more parental nucleic acids (e.g., with a nuclelease, or via chemical A collection of "fragmented nucleic acids" is a collection of nucleic acids

nucleic acids are hybridized and elongated).

30 substrate for recombination (i.e., when one or more oligonucleotide(s) corresponding to the sequences is an actual substrate for recombination, and in some cases, neither sequence is a

as a cross-over oligonucleotide. For indirect recombination, no more than one of the are "indirectly recombinant" when the sequences are recombinant using an intermediate such recombined when both of the nucleic acids are substrates for recombination. Two sequences nucleic acids are combined in a progeny nucleic acid. Two sequences are "directly"

25 Two nucleic acids are "recombined" when sequences from each of the two

termini of the nucleic acid.

polymerase (e.g., a DNA polymerase), e.g., a polymerase which adds sequences at the 3' molecules) are incorporated into the nucleic acid. Most commonly, this is performed with a Nucleic acids are "elongated" when additional nucleotides (or other analogous

20 derived, by natural or artificial manipulation, from the other.

sequences, or when one nucleic acid is a subsequence of the other, or when one sequence is

Two nucleic acids "complement" when they have the same or complementary

Elsiever, New York, as well as in Ausubel, *supra*.

2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays,"

Biotechnology and Molecular Biology-Hybridization with Nucleic Acid Probes part I chapter

15 hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in* hydrogen bonding, solvent exclusion, base stacking and the like. An extensive guide to the acids hybridize due to a variety of well characterized physico-chemical forces, such as Nucleic acids "hybridize" when they associate, typically in solution. Nucleic

10 homologous.

different members of a selected set of nucleic acids, which are optionally homologous or non-

A "cross-over" oligonucleotide has regions of sequence identity to at least two homologous nucleic acids (e.g., about 2, 3, 5, 10, 50 or more member types).

5 each having regions of sequence identity to at least one member of the selected set of Most commonly, the family gene shuffling oligonucleotides include multiple member types, apphcations, 80%, 90% or 100% of the full-length of each of the homologous nucleic acids). Generally 50% or more, typically 60% or more, more typically 70% or more, and in some oligonucleotides of the set collectively correspond to e.g., 25% or more, often 35% or more, correspond over a substantial portion of the length of the homologous nucleic acids (e.g., the

90% or more, typically 60% or more, more typically 70% or more, and in some

WO 00/42561 PCT/US00/01203

cleavage), or by producing subsequences of the parental sequences in any other manner, such as partial chain elongation of a complementary nucleic acid.

A "full-length protein" is a protein having substantially the same sequence domains as a corresponding protein encoded by a natural gene. The protein can have 5 modified sequences relative to the corresponding naturally encoded gene (e.g., due to recombination and selection), but is at least 95% as long as the naturally encoded gene.

A "DNase enzyme" is an enzyme such as DNase I which catalyzes cleavage of a DNA, *in vitro* or *in vivo*. A wide variety of DNase enzymes are well known and described, e.g., in Sambrook, Berger and Ausubel (*all supra*) and many are commercially 10 available.

A "nucleic acid domain" is a nucleic acid region or subsequence. The domain can be conserved or not conserved between a plurality of homologous nucleic acids. Typically, a domain is delineated by comparison between two or more sequences, i.e., a 15 region of sequence diversity between sequences is a "sequence diversity domain," while a region of similarity is a "sequence similarity domain." Domain switching" refers to the ability to switch one nucleic acid region from one nucleic acid with a second domain from a second nucleic acid.

A region of "high sequence similarity" refers to a region that is 90% or more identical to a second selected region when aligned for maximal correspondence (e.g., 20 manually or using the common program BLAST set to default parameters). A region of "low sequence similarity" is 60% or less identical, more preferably, 40% or less identical to a second selected region, when aligned for maximal correspondence (e.g., manually or using BLAST set with default parameters).

A "PCR amplicon" is a nucleic acid made using the polymerase chain reaction (PCR). Typically, the nucleic acid is a copy of a selected nucleic acid. A "PCR primer" is a 25 nucleic acid which hybridizes to a template nucleic acid and permits chain elongation using a thermostable polymerase under appropriate reaction conditions.

A "library of oligonucleotides" is a set of oligonucleotides. The set can be pooled, or can be individually accessible. Oligonucleotides can be DNA, RNA or 30 combinations of RNA and DNA (e.g., chimeroplasts).

DETAILED DISCUSSION OF THE INVENTION

The present invention relates to improved formats for nucleic acid shuffling. In particular, by using selected oligonucleotide sets as substrates for recombination and/or gene synthesis, it is possible to dramatically speed the shuffling process. Moreover, it is

In general, the methods herein provide for "oligonucleotide mediated shuffling" in which oligonucleotides corresponding to a family of related homologous nucleic acids which are recombinated to produce selectable nucleic acids. The technique can be used to recombine homologous or even non-homologous nucleic acid sequences. When to recombine homologous nucleic acids, sets of overlapping family gene shuffling recombining homologous nucleic acids, sets of overlapping family gene shuffling

However, one advantage of the present invention is the ability to recombine several regions of diversity among homologous nucleic acids, even without the homologous nucleic acids, or cleaved fragments thereof, being present in the recombination mixture. Resulting shuffled nucleic acids can include regions of diversity from different nucleic acids, providing for the ability to combine different diversity domains in a single nucleic acid. This provides a very powerful method of accessing natural sequence diversity.

reconstituted into a full length sequence—an advantage in the event that one or more synthesized oligo in the synthesized set is defective.

In brief, a family of homologous nucleic acid sequences are first aligned, e.g., using available computer software to select regions of identity/similarity and regions of diversity. A plurality (e.g., 2, 5, 10, 20, 50, 75, or 100 or more) of oligonucleotides corresponding to at least one region of diversity (and ordinarily at least one region of similarity) are synthesized. These oligonucleotides can be shuffled directly, or can be combined with one or more of the family of nucleic acids. This oligonucleotide-based recombination of related nucleic acids can be combined with a number of available standard shuffling methods. For example, there are several procedures now available for shuffling homologous nucleic acids, such as by digesting the nucleic acids with a DNase, permitting recombination to occur and then

possible to use oligonucleotide intermediates to indirectly recombine nucleic acids which could not otherwise be recombined. Direct access to physical nucleic acids corresponds to sequences to be otherwise combined. Directly combined sequences can be recombinably combined, as the sequences can be recombinably combined indirectly through oligonucleotide intermediates.

oligonucleotides (which are derived, e.g., by comparison of homologous nucleic acids and synthesis of oligonucleotide fragments) are hybridized and elongated (e.g., by reassembly PCR), providing a population of recombined nucleic acids, which can be selected for a desired trait or property. Typically, the set of overlapping family shuffling gene 5 oligonucleotides include a plurality of oligonucleotide member types which have consensus region subsequences derived from a plurality of homologous target nucleic acids.

Typically, family gene shuffling oligonucleotide are provided by aligning homologous nucleic acid sequences to select conserved regions of sequence identity and regions of sequence diversity. A plurality of family gene shuffling oligonucleotides are 10 synthesized (serially or in parallel) which correspond to at least one region of sequence diversity.

Sets of fragments, or subsets of fragments used in oligonucleotide shuffling approaches can be partially provided by cleaving one or more homologous nucleic acids (e.g., with a DNase), as well as by synthesizing a set of oligonucleotides corresponding to a 15 plurality of regions of at least one nucleic acid (typically oligonucleotides corresponding to a partial or full-length nucleic acid are provided as members of the set of nucleic acid "fragments," a term which encompasses both cleavage fragments and synthesized oligonucleotides): In the shuffling procedures herein, these cleavage fragments can be used in conjunction with family gene shuffling oligonucleotides, e.g., in one or more 20 recombination reaction to produce recombinant nucleic acids.

The following provides details and examples regarding sequence alignment, oligonucleotide construction and library generation, shuffling procedures and other aspects of the present invention.

ALIGNING HOMOLOGOUS NUCLEIC ACID SEQUENCES TO SELECT CONSERVED 25 REGIONS OF SEQUENCE IDENTITY AND REGIONS OF SEQUENCE DIVERSITY

In one aspect, the invention provides for alignment of nucleic acid sequences to determine regions of sequence identity or similarity and regions of diversity. The set of overlapping family shuffling gene oligonucleotides can comprise a plurality of oligonucleotide member types which comprise consensus region subsequences derived from 30 a plurality of homologous target nucleic acids. These consensus region subsequences are determined by aligning homologous nucleic acids and identifying regions of identity or similarity.

parameters are designated. The sequence comparison algorithm then calculates the percent sequence coordinates are designated, if necessary, and sequence algorithm program sequence acts as a reference sequence to which test sequences are compared. When using a sequence acts as a reference sequence and homology determination, typically one

30 sequences are typically considered to be homologous. For sequence comparison and homology determination, typically one sequence comparison algorithms or by visual inspection. Such "substantially identical" sequence comparison algorithms correspondence, as measured using one of the following compared and aligned for maximum correspondence, as measured using one of the following preferably 80%, most preferably 90-95% nucleotide or amino acid residue identity, when polypeptides refers to two or more sequences or subsequences that have at least about 50%, The phrase "substantially identical," in the context of two nucleic acids or

25 persons of skill) or by visual inspection. The sequence comparison algorithms described below (or other algorithms available to same, when compared and aligned for maximum correspondence, as measured using one of are the same or have a specified percentage of amino acid residues or nucleotides that are the nucleic acid or polyptide sequences, refer to two or more sequences or subsequences that The terms "identical" or percent "identity," in the context of two or more

30 companies specializing in genomic information generation and/or storage. Sequence databases are available on the internet or on a contract basis from a variety of Genbank®, Entrez®, EMBL, DBJ and those provided by the NCB. Many additional available or proprietary nucleic acid databases. Public database/ search services include regions of nucleic acid similarity, e.g., for sequences available from any of the public. Typically, the invention compares first aligning identical nucleic acids, or

10 Alignment of homologous nucleic acids and/or mutations.

5 further screened or further diversified by any available method, including recombination otherwise available. Genes synthesized by these recombination methods are optionally synthesized and used in the various recombination protocols noted herein or which are oligonucleotides corresponding to the plurality of domains of sequence diversity are

diversity are selected. The plurality of regions of sequence diversity provide a plurality of least one conserved region of sequence identity and a plurality of regions of sequence domains of sequence diversity. Typically, a plurality of family gene shuffled

sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally*, Ausubel *et al.*, *infra*).

One example algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W , T , and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, $M=5$, $N=-4$, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

many others.

ExpressGen Inc. (www.expressgen.com), Operon Technologies Inc. (Alameda, CA) and (mcrc@oligos.com), The Great American Gene Company (http://www.genco.com), variety of commercial sources, such as The Midland Certified Reagent Company.

30 Moreover, essentially any nucleic acid can be custom ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company.

useful.

25 Nucleotide synthesis approaches (e.g., tri-nucleotide synthesis), as discussed, *supra*, are also equipment is commercially available for automated oligonucleotide synthesis. Multi-phosphoramidite triester method described by Beauchage and Caruthers (1981), *Terahedron Lett.*, 22(20):1859-1862, e.g., using an automated synthesizer, as described in Needham-Vandewater *et al.* (1984) *Nucleic Acids Res.*, 12:6159-6168. A wide variety of oligonucleotides, are typically synthesized chemically according to the solid phase reconstruction/reassembly methods, or to provide sets of family gene shuffling Oligonucleotides, e.g., whether for use in *in vitro* amplification, gene parallel oligonucleotide synthesis protocols.

20 Typically sets of family gene shuffling oligonucleotides are produced, e.g., by sequential or shuffling oligonucleotides, e.g., corresponding to at least one region of sequence diversity. In one aspect, the invention comprises synthesizing a plurality of family gene parallel oligonucleotide synthesis

15

POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Seifionov *et al.*, attorney sequence alignment programs include, e.g., PILFUP.

10 A number of additional sequence alignment protocols can be found, e.g., in "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES AND

5 ALTSCHUL (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence (and, therefore, likely homologous) if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001. Other available sequence alignment programs include, e.g., PILFUP.

100

20

25

30

35

40

45

50

55

60

65

70

75

80

85

90

95

100

105

110

115

120

125

130

135

140

145

150

155

160

165

170

175

180

185

190

195

200

205

210

215

220

225

230

235

240

245

250

255

260

265

270

275

280

285

290

295

300

305

310

315

320

325

330

335

340

345

350

355

360

365

370

375

380

385

390

395

400

405

410

415

420

425

430

435

440

445

450

455

460

465

470

475

480

485

490

495

500

505

510

515

520

525

530

535

540

545

550

555

560

565

570

575

580

585

590

595

600

605

610

615

620

625

630

635

640

645

650

655

660

665

670

675

680

685

690

695

700

705

710

715

720

725

730

735

740

745

750

755

760

765

770

775

780

785

790

795

800

805

810

815

820

825

830

835

840

845

850

855

860

865

870

875

880

885

890

895

900

905

910

915

920

925

930

935

940

945

950

955

960

965

970

975

980

985

990

995

1000

1005

1010

1015

1020

1025

1030

1035

1040

1045

1050

1055

1060

1065

1070

1075

1080

1085

1090

1095

1100

1105

1110

1115

1120

1125

1130

1135

1140

1145

1150

1155

1160

1165

1170

1175

1180

1185

1190

1195

1200

1205

1210

1215

1220

1225

1230

1235

1240

1245

1250

1255

1260

1265

1270

1275

1280

1285

1290

1295

1300

1305

1310

1315

1320

1325

1330

1335

1340

1345

1350

1355

1360

1365

1370

1375

1380

1385

1390

1395

1400

1405

1410

1415

1420

1425

1430

1435

1440

1445

1450

1455

1460

1465

1470

1475

1480

1485

1490

1495

1500

1505

1510

1515

1520

1525

1530

1535

1540

1545

1550

1555

1560

1565

1570

1575

1580

1585

1590

1595

1600

1605

1610

1615

1620

1625

1630

1635

1640

1645

1650

1655

1660

1665

1670

1675

1680

1685

1690

1695

1700

1705

1710

1715

1720

1725

1730

1735

1740

1745

1750

1755

1760

1765

1770

1775

1780

1785

1790

1795

1800

1805

1810

1815

1820

1825

1830

1835

1840

1845

1850

1855

1860

1865

1870

1875

1880

1885

1890

1895

1900

1905

1910

1915

1920

1925

1930

1935

1940

1945

1950

1955

1960

1965

1970

1975

1980

1985

1990

1995

2000

2005

2010

2015

2020

2025

2030

2035

2040

2045

2050

2055

2060

2065

2070

2075

2080

2085

2090

2095

2100

2105

2110

2115

2120

2125

2130

2135

2140

2145

2150

2155

2160

2165

2170

2175

2180

2185

2190

2195

2200

2205

2210

2215

2220

2225

2230

2235

2240

2245

2250

2255

2260

2265

2270

2275

2280

2285

2290

2295

2300

2305

2310

2315

2320

2325

2330

2335

2340

2345

2350

2355

2360

2365

2370

2375

2380

2385

2390

2395

2400

2405

2410

2415

2420

2425

2430

2435

2440

2445

2450

2455

2460

2465

2470

2475

2480

2485

2490

2495

2500

2505

2510

2515

2520

2525

2530

2535

2540

2545

2550

2555

2560

2565

2570

2575

2580

2585

2590

2595

2600

2605

2610

2615

2620

2625

2630

2635

2640

2645

2650

2655

2660

2665

2670

2675

2680

2685

2690

2695

2700

2705

2710

2715

2720

2725

2730

2735

2740

2745

2750

2755

2760

2765

2770

2775

2780

2785

2790

2795

2800

2805

2810

2815

2820

2825

2830

2835

2840

2845

2850

2855

2860

2865

2870

2875

2880

2885

2890

2895

2900

2905

2910

2915

2920

2925

2930

2935

2940

2945

2950

2955

2960

2965

2970

2975

2980

2985

2990

2995

3000

3005

3010

3015

3020

3025

3030

3035

3040

3045

3050

3055

3060

3065

3070

3075

3080

3085

3090

3095

3100

3105

3110

3115

3120

3125

3130

3135

3140

3145

3150

3155

3160

3165

3170

3175

3180

3185

3190

3195

3200

3205

3210

3215

3220

3225

3230

3235

3240

3245

3250

3255

3260

3265

3270

3275

3280

3285

3290

3295

3300

3305

3310

3315

3320

3325

3330

3335

3340

3345

3350

3355

3360

3365

3370

3375

3380

3385

3390

3395

3400

3405

3410

3415

3420

3425

3430

3435

3440

3445

3450

3455

3460

3465

3470

3475

3480

3485

3490

3495

3500

3505

3510

3515

3520

3525

3530

3535

3540

3545

3550

3555

3560

3565

3570

3575

3580

3585

3590

3595

3600

3605

3610

3615

3620

3625

3630

3635

3640

3645

3650

3655

3660

3665

3670

3675

3680

3685

3690

3695

3700

3705

3710

3715

3720

3725

3730

3735

3740

3745

3750

3755

3760

3765

3770

3775

3780

3785

3790

3795

3800

3805

3810

3815

3820

3825

3830

3835

3840

3845

3850

3855

3860

3865

3870

3875

3880

3885

3890

3895

3900

3905

3910

3915

3920

3925

3930

3935

3940

3945

3950

3955

3960

3965

3970

3975

3980

3985

3990

3995

4000

4005

4010

4015

4020

4025

4030

4035

4040

4045

4050

4055

4060

4065

4070

4075

4080

4085

4090

4095

4100

4105

4110

4115

4120

4125

4130

4135

4140

4145

4150

4155

4160

4165

4170

4175

4180

4185

4190

4195

4200

4205

4210

4215

4220

4225

4230

4235

4240

4245

4250

4255

4260

4265

4270

4275

4280

4285

4290

4295

4300

4305

4310

4315

4320

4325

4330

4335

4340

4345

4350

4355

4360

4365

4370

4375

4380

4385

4390

4395

4400

4405

4410

4415

4420

4425

4430

4435

4440

4445

4450

4455

4460

4465

4470

4475

4480

4485

4490

4495

4500

4505

4510

4515

4520

4525

4530

4535

4540

4545

4550

4555

4560

4565

4570

4575

4580

4585

4590

4595

4600

4605

4610

4615

4620

4625

4630

4635

4640

4645

4650

4655

4660

4665

4670

4675

4680

4685

4690

4695

4700

4705

4710

4715

4720

4725

4730

4735

4740

4745

4750

4755

4760

4765

4770

4775

4780

4785

4790

4795

4800

4805

4810

4815

4820

4825

4830

4835

4840

4845

4850

4855

4860

4865

4870

4875

4880

4885

4890

4895

4900

4905

4910

4915

4920

4925

4930

4935

4940

4945

4950

4955

4960

4965

4970

4975

4980

4985

4990

4995

5000

5005

5010

5015

5020

5025

5030

5035

5040

5045

5050

5055

5060

5065

5070

5075

5080

5085

5090

5095

5100

5105

5110

5115

5120

5125

5130

5135

5140

5145

5150

5155

5160

5165

5170

5175

5180

5185

5190

5195

5200

5205

5210

5215

5220

5225

5230

5235

5240

5245

5250

5255

5260

5265

5270

5275

5280

5285

5290

5295

5300

5305

5310

5315

5320

5325

5330

5335

5340

5345

5350

5355

5360

5365

5370

5375

5380

5385

5390

5395

5400

5405

5410

5415

5420

5425

5430

5435

5440

5445

5450

5455

5460

5465

5470

5475

5480

5485

5490

5495

5500

5505

5510

5515

5520

5525

5530

5535

5540

5545

5550

5555

5560

5565

5570

5575

5580

5585

5590

5595

5600

5605

5610

5615

5620

5625

5630

5635

5640

5645

5650

5655

5660

5665

5670

5675

5680

5685

5690

5695

5700

5705

5710

5715

5720

5725

5730

5735

5740

5745

5750

5755

5760

5765

5770

5775

5780

5785

5790

5795

5800

5805

5810

5815

5820

5825

5830

5835

5840

5845

5850

5855

5860

5865

5870

5875

5880

5885

5890

5895

5900

5905

5910

5915

5920

5925

5930

5935

5940

5945

5950

5955

5960

5965

5970

5975

5980

5985

5990

5995

6000

6005

6010

6015

Synthetic Library Assembly

Libraries of family gene shuffling oligonucleotides are provided. For example, homologous genes of interest are aligned using a sequence alignment program such as BLAST, as described above. Nucleotides corresponding to amino acid variations between the homologs are noted. Oligos for synthetic gene shuffling are designed which comprise one (or more) nucleotide difference to any of the aligned homologous sequences, i.e., oligos are designed that are identical to a first nucleic acid, but which incorporate a residue at a position which corresponds to a residue of a nucleic acids homologous, but not identical to the first nucleic acid.

Preferably, all of the oligonucleotides of a selected length (e.g., about 20, 30, 40, 50, 60, 70, 80, 90, or 100 or more nucleotides) which incorporate all possible nucleic acid variants are made. This includes X oligonucleotides per X sequence variations, where X is the number of different sequences at a locus. The X oligonucleotides are largely identical in sequence, except for the nucleotide(s) representing the variant nucleotide(s). Because of this similarity, it can be advantageous to utilize parallel or pooled synthesis strategies in which a single synthesis reaction or set of reagents is used to make common portions of each oligonucleotide. This can be performed e.g., by well-known solid-phase nucleic acid synthesis techniques, or, e.g., utilizing array-based oligonucleotide synthetic methods (see e.g., Fodor et al. (1991) *Science*, 251: 767- 777; Fodor (1997) "Genes, Chips and the Human Genome" *FASEB Journal*, 11:121-121; Fodor (1997) "Massively Parallel Genomics" *Science*, 277:393-395; and Chee et al. (1996) "Accessing Genetic Information with High-Density DNA Arrays" *Science* 274:610-614). Additional oligonucleotide synthetic strategies are found, e.g., in "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES AND POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov et al., attorney docket number 02-289-3US, filed herewith.

In one aspect, oligonucleotides are chosen so that only encoded amino acid alterations are considered in the synthesis strategy. In this strategy, after aligning a family of homologous nucleic acids, family shuffling oligos are synthesized to be degenerate only at those positions where a base change results in an alteration in an encoded polypeptide sequence. This has the advantage of requiring fewer degenerate oligonucleotides to achieve the same degree of diversity in encoded products, thereby simplifying the synthesis of the set of family gene shuffling oligonucleotides.

289-3US, filed herewith). As above, family shuffling oligos are designed to incorporate HAVING DESIRED CHARACTERISTICS" by Seltznerov et al., attorney docket number 02- FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES AND POLYPEPTIDES manually, e.g., on a printout of the sequences or sequence alignments. See also, "METHODS 30 system comprising a computer running appropriate sequence alignment software, or described above. Amino acid variations are noted and/or marked (e.g., in an integrated homologous set of genes). In one aspect, all of the nucleic acid to be shuffled are aligned as typical shuffling mixture; e.g., a mixture of DNA fragments of one or more gene(s) from a family oligonucleotides can also be used to vary the nucleic acids present in a

25 **Library Splicing**
 hybridizing and elongating those sequences again. This cycle can be repeated for any desired number of times. The cycle is repeated e.g., from about 2 to about 100 times. denaturing the resulting elongated double-stranded nucleic acid sequences and then hybridizing and elongating those sequences again. This cycle can be repeated for any desired number of times. The cycle is repeated e.g., from about 2 to about 100 times. noted elsewhere in this disclosure, certain embodiments of the invention can involve 20 any additional nucleic acids which form a recombination mixture) prior to, simultaneously with, or after hybridization of the oligonucleotides and other recombination components. As temperature. The polymerase can be added to nucleic acid fragments (oligonucleotides plus Klenow, Taq or polymerases can be used with a hybridization temperature of below room temperature and e.g., about 65°C. If the areas of identity are small, between about room temperature and, e.g., about 65°C. If the areas of identity are small, or other high-temperature polymerase can be used with a hybridization temperature of 15 DNTPs (i.e., dATP, dCTP, dGTP and dTTP). If regions of sequence identity are large, Taq are incubated in the presence of a nucleic acid polymerase, e.g., Taq, Klenow, or the like, and For example, during oligonucleotide elongation, hybridized oligonucleotides 20 incorporate incorporation of desirable modifications.

10 bases of sequence identity to either side of a region of variance to ensure reasonably efficient hybridization and assembly. Oligonucleotides can be incubated together and reassembled using any of a variety of polymerase-mediated reassembly methods, e.g., as described herein and as known to one of skill. Selected oligonucleotides can be "spiked" in the recombination mixture at any selected concentration, thus causing preferential incorporation of desirable modifications.

15 In synthesis strategies in general, the oligonucleotides have at least about 10 identical bases (e.g., 5, 6, 7, 8, or 9) and can, of course, have larger regions of identity (e.g., 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 50, or more). During gene assembly, oligonucleotides can be incubated together and

some or all of the amino acid variations coded by the natural sequence diversity for the aligned nucleic acids. One or more nucleic acids corresponding to the homologous set of aligned nucleic acids are cleaved (e.g., using a DNase, or by chemical cleavage). Family shuffling oligos are spiked into the mixture of cleaved nucleic acids, which are then 5 recombined and reassembled into full-length sequences using standard techniques.

To determine the extent of oligonucleotide incorporation, any approach which distinguishes similar nucleic acids can be used. For example, the reassembled nucleic acids can be cloned and sequenced, or amplified (in vitro or by cloning, e.g., into a standard cloning vector) and cleaved with a restriction enzyme which specifically recognizes a 10 particular polymorphic sequence present in the family shuffling oligos, but not present in the same position in the original cleaved nucleic acid(s).

In another embodiment, oligonucleotides are selected which incorporate one or more sequence variation corresponding to an amino acid polymorphism, but which eliminate polymorphic nucleotide variations between nucleic acid sequences which 15 correspond to silent substitutions. One advantage of this strategy is that the elimination of silent substitutions can make a given sequence more similar to a given substrate for recombination (e.g., a selected target nucleic acid). This increased similarity permits nucleic acid recombination among sequences which might otherwise be too diverse for efficient recombination.

For example, a selected nucleic acid can be PCR amplified using standard 20 methods. The selected nucleic acid is cleaved and mixed with a library of family gene shuffling oligonucleotides which are rendered as similar as possible to the corresponding sequences of the selected nucleic acid by making the oligonucleotides include the same silent substitution set found in the selected nucleic acid. The oligonucleotides are spiked at a 25 selected concentration into the cleavage mixture, which is then reassembled into full-length sequences. The quality of the resulting library (e.g., frequency at which the oligos are incorporated into the reassembled sequences) is checked, as noted above, by cloning (or otherwise amplifying) and sequencing and/or restriction digesting the reassembled sequences.

PCR elongation strategies can also be used to make libraries using different 30 molar ratios of oligonucleotides in the recombination mixtures (see also, e.g., WO 97/20078, WO 98/42832 and WO 98/01581).

In addition, recombinant nucleic acids made using family shuffling are, optionally, reiterative. Standard recombination can be used in conjunction with oligonucleotides can be cleaved and shuffled using standard recombination methods, which are, optionally, reiterative. Standard recombination can be used in conjunction with oligonucleotide shuffling and either or both steps are optionally reiteratively repeated. One useful example of iterative shuffling by oligonucleotide mediated recombination of family oligonucleotides occurs when extremely fine grain shuffling is desired. For example, small genes encoding small protein such as defensins (antifungal proteins of about 50 amino acids) Ff40 (an antifungal protein family of about 28 amino acids), peptide antibiotics, peptide insecticidal proteins, peptide hormones, many cytokines because the recombination often occurs with a frequency that is roughly the same as the size and many other small proteins, are difficult to recombine by standard recombination methods, because the recombination often occurs with a frequency that is roughly the same as the size of the gene to be recombined, limiting the diversity resulting from recombination. In contrast, oligonucleotide-mediated recombination methods can recombine essentially any region of diversity in any set of sequences, with recombination events (e.g., crossovers) occurring at any selected base-pair.

Thus, libraries of sequences prepared by recursive oligonucleotide mediated recombination are optionally screened and selected for a desired property, and improved (or otherwise desirable) clones are sequenced (or otherwise deconvoluted, e.g., by real time PCR being iteratively repeated to generate additional libraries of nucleic acids. Thus, additional recombination rounds are performed either by standard fragmentation-based recombination methods, or by sequencing positive clones, designing appropriate family shuffling

oligonucleotides and performing a second round of recombination/selection to produce an oligonucleotide library.

In one aspect, the present invention provides iterative oligonucleotide-mediated recombination methods. These formats can be combined with standard mediated recombination methods, also, optionally, in an iterative format. In particular, recombinant nucleic acids produced by oligonucleotide-mediated recombination can be screened for activity and sequenced. The sequenced recombinant nucleic acids are aligned and regions of identity and diversity are identified. Family shuffling oligonucleotides are then selected for recombination of the sequenced recombinant nucleic acids. This process of screening, sequencing active recombinant nucleic acids and combining the active recombinant nucleic acids can be iteratively repeated until a molecule with a desired property is obtained.

Terative Oligonucleotide Formats

additional library (which can be recombined as described). In addition, libraries made from different recombination rounds can also be recombined, either by sequencing/oligonucleotide recombination or by standard recombination methods.

Crossover PCR Shuffling

5 In one aspect, the present invention provides for shuffling of distantly related or even non-homologous sequences. In this embodiment, PCR crossover oligonucleotides are designed with a first region derived from a first nucleic acid and a second region corresponding to a second nucleic acid. Additional oligos are designed which correspond to either the first or second nucleic acid, and which have sequences that are complementary (or 10 identical) to the crossover oligos. By recombining these oligos (i.e., hybridizing them and then elongating the hybridized oligonucleotides in successive polymerase-mediated elongation reactions), a substrate is provided which can recombine with either the first or second nucleic acid, and which will, at the same time, incorporate sequences from the other nucleic acid.

15 **IN VIVO OLIGONUCLEOTIDE RECOMBINATION UTILIZING FAMILY SHUFFLING CHIMERAPLASTS**

Chimeraplasts are synthetic RNA-DNA hybrid molecules which have been used for "genetic surgery" in which one or a few bases in a genomic DNA are changed by recombination with the chimeric molecule. The chimeraplasts are chimeric nucleic acids 20 composed of contiguous stretches of RNA and DNA residues in a duplex conformation with double hairpin caps on the ends of the molecules (Yoon et al. (1996) PNAS 93:2071-2076). The RNA-DNA sequence is designed to align with the sequence of a locus to be altered by recombination with the chimeraplast, with the chimeraplast having the desired change in base sequence for the locus. The host cell repair machinery converts the host cell sequence to that 25 of the chimeraplast. For brief reviews of the technique see, Bartlett (1998) Nature Biotechnology 16:1312; Strauss (1998) Nature Medicine 4:274-275.

This strategy has been used for targeted correction of a point mutation in the gene for human liver/kidney/bone alkaline phosphatase encoded on an episomal DNA in mammalian cells (Yoon, *id.*). The strategy was also used for correction of the mutation 30 responsible for sickle cell anemia in genomic DNA in lymphoblastoid cells (Cole-Strauss et al. (1996) Science 1386-1389). Alexeev and Yoon (1998) Nature Biotechnology 1343-1346 describe the use of a hybrid RNA-DNA oligonucleotide (an "RDO") to make a point correction in the mouse tyrosinase gene, resulting in correction of an albino mutation in

or more homologous sequences. Thus, the effect of transducing the cell population with the interest within the cells is converted at one or more positions to a sequence derived from one library of homologous chimeraplast sequences into a population of cells, the target gene of a subsequence of a gene into a subsequence found in an homologous gene. By transducing a 30 incorporation of the chimeraplasts into one or more target sequences, results in conversion of acids. That is, the chimeraplasts each contain one or a few nucleotides which, following to regions of sequence diversity observed following an alignment of homologous nucleic For example, chimeraplasts are synthesized with sequences which correspond

25 described.

different set of chimeraplast oligonucleotides, followed by selection/screening assays as members are subjected to an additional round of chimeraplast recombination with the same or library is then screened and selected as described herein. Optionally, the selected library creating a library of cells with at least one region of diversity at a target gene of interest. The the chimeraplasts combine with a sequence of interest in the genome of the cells, thereby forth in Figure 1, the set of oligonucleotides is transduced into cells (e.g., plant cells), where sequence variation for a particular subsequence) found in the target gene of interest. As set target gene of interest, and regions of diversity corresponding to the diversity (i.e., the

20 The set of family shuffling chimeraplasts each include regions of identity to a

15 references noted above.

standard nucleotides such as 2'-O methylated RNA nucleotides). The oligos have a "dual phosphoramidite coupling chemistry (the nucleotides utilized optionally include non-hairpin" structure (e.g., having a T loop at the ends of the structure) as set forth in the references noted above, DNA-RNA oligos are synthesized according to standard forth herein, to additionally include structural chimeraplast features. For example, in the are chimeraplasts. In this embodiment, family shuffling oligonucleotides are made as set forth herein, to additionally include structural chimeraplast features. For example, in the 10 In one aspect of the present invention, the family shuffling oligonucleotides are chimeric RNA-DNA oligonucleotide.

targeted nucleotide exchange in the alkaline phosphatase gene of Hu-H-7 cells mediated by a RNA-DNA oligonucleotide. Kren et al. (1997) *Hepatology* 25(6):1462-1468 describe targeted gene conversion in a mammalian CD34⁺-enriched cell population using a chimeric chimeric RNA/DNA oligonucleotides. Xiang et al (1997) *J. Mol. Med.* 75:829-835 describe Medicine 4(3):285-290 describe in vivo site-directed mutagenesis of the factor IX gene by mouse cells and production of black pigmentation by the cells. Kren et al. (1998) *Nature* 5

chimeraplast library is to create a library of target genes corresponding to the sequence diversity found in genes homologous to the target sequence.

Chimeraplasts can also be similarly used to convert the target gene at selected positions with non-homologous sequence choices, e.g., where structural or other information suggests the desirability of such a conversion. In this embodiment, the chimeraplasts include sequences corresponding to non-homologous sequence substitutions.

Optionally, the chimeraplasts, or a co-transfected DNA, can incorporate sequence tags, selectable markers, or other structural features to permit selection or recovery of cells in which the target gene has recombined with the chimeraplast. For example, a co-transfected DNA can include a marker such as drug resistance, or expression of a detectable marker (e.g., *Lac Z*, or green fluorescent protein).

In addition, sequences in the chimeraplast can be used as purification or amplification tags. For example, a portion of the chimeraplast can be complementary to a PCR primer. In this embodiment, PCR primers are used to synthesize recombinant genes from the cells of the library. Similarly, PCR primers can bracket regions of interest, including regions in which recombination between a chimeraplast and a standard DNA occurs. Other PCR, restriction enzyme digestion and/or cloning strategies which result in the isolation of nucleic acids resulting from recombination between the chimeraplast can also be used to recover the recombined nucleic acid, which is optionally recombined with additional nucleic acids. Reiterative cycles of chimeraplast-mediated recombination, recovery of recombinant nucleic acids and recombination of the recovered nucleic acids can be performed using standard recombination methods. Selection cycles can be performed after any recombination event to select for desirable nucleic acids, or, alternatively, several rounds of recombination can be performed prior to performing a selection step.

25 LIBRARIES OF CHIMERAPLASTS AND OTHER GENE RECOMBINATION
VEHICLES

As noted above, chimeraplasts are generally useful structures for modification of nucleotide sequences in target genes, *in vivo*. Accordingly, structures which optimize chimeraplast activity are desirable. Thus, in addition to the *use* of chimeraplasts in *in vitro* and *in vivo* recombination formats as noted, the present invention also provides for the optimization of chimeraplast activity *in vitro* and *in vivo*, as well as for a number of related libraries and other compositions.

The library which is produced by these methods is screened for increased recombination rates as noted above. Library members which are identified as having increased rates of recombination are optionally themselves recombined to produce libraries of increased rates of recombination. Library members which are identified as having increased rates of recombination are optionally themselves recombined to produce libraries of increased rates of recombination. Recombination is ordinarily performed by assessing the recombinant chimeraplasts. Recombination is ordinarily performed by assessing the sequences of the members which initially display increased recombination rates, followed by

which provide for production of heterogeneous sequences as described herein. For example, synthetic strategies utilizing chimeraplasts which are largely identical in sequence, except for variant nucleotide(s) are produced to simplify synthetic strategies. Because of this similarity, parallel or pooled synthesis strategies can be used in which a single synthesis reaction or set of reagents is used to make common portions of each oligonucleotide. This can be performed e.g., by well-known solid-phase nucleic acid synthesis techniques, e.g., in a commercially available oligonucleotide synthesizer, or, e.g., by utilizing array-based oligonucleotide synthesis methods (see e.g., Fodor et al. (1991) *Science*, 251: 767-777; Fodor (1997) "Genes, Chips and the Human Genome" *FASBB Science*, 11:121-121; Fodor (1997) "Massively Parallel Genomics" *Science*, 277:393-395; and Chee et al. (1996) "Accessing Genetic Information with High-Density DNA Arrays" *Science* 274:610-614). According to the present invention is a library of chimeraplasts produced by these methods, i.e., a library of chimeraplasts which share common sequence elements, including e.g., a common marker, as well as regions of difference, e.g., different sequences in the hairpin regions of the molecule.

In particular, a marker can be incorporated into a library of related chimeraplasts. The marker is placed between the ends of the chimeraplast in the region of the molecule which is incorporated into a target nucleic acid. For example, the marker can cause a detectable phenotype effect in a cell in which recombination occurs, or the marker can simply lead to a change in the target sequence which can be detected by standard nucleic acid sequence detection techniques (e.g., PCR amplification of the sequence or of a flanking sequence, LCR, restriction enzyme digestion of a sequence created by a recombination event, binding of the recombinant nucleic acid to an array (e.g., a gene chip), and/or sequencing of the recombinant nucleic acid, etc.). Ordinarily, the regions of sequence difference are determined to provide an indication of which sequences have increased recombination rates. The library of related chimeraplasts includes chimeraplasts with regions of sequence difference in the T loop hairpin regions and in the region between the T loop hairpin regions of the market. This divergence can be produced by synthetic strategies hairpin regions flanking the market. This divergence can be produced by synthetic strategies

synthesis of chimeraplasts which display structural similarity to at least two of these members. This process can be iteratively repeated to create new "recombinant" chimeraplasts with increased recombination activity, as well as libraries of such chimeraplasts.

5 Other recombination molecules can similarly be produced by these methods. For example, Cre-Lox sites, Chi sites and other recombination facilitating sequences in cell transduction/transformation vectors are varied and selected in the same manner as noted above. Where the sequences are simple DNA sequences, they can be recombined either by the synthetic methods noted herein, and/or by standard DNA shuffling methods.

10 **CODON-VARIED OLIGONUCLEOTIDES**

Codon-varied oligonucleotides are oligonucleotides, similar in sequence but with one or more base variations, where the variations correspond to at least one encoded amino acid difference. They can be synthesized utilizing tri-nucleotide, i.e., codon-based phosphoramidite coupling chemistry, in which tri-nucleotide phosphoramidites representing 15 codons for all 20 amino acids are used to introduce entire codons into oligonucleotide sequences synthesized by this solid-phase technique. Preferably, all of the oligonucleotides of a selected length (e.g., about 20, 30, 40, 50, 60, 70, 80, 90, or 100 or more nucleotides) which incorporate the chosen nucleic acid sequences are synthesized. In the present invention, 20 codon-varied oligonucleotide sequences can be based upon sequences from a selected set of homologous nucleic acids.

The synthesis of tri-nucleotide phosphoramidites, their subsequent use in oligonucleotide synthesis, and related issues are described in, e.g., Virnekäs, B., *et al.*, (1994) *Nucleic Acids Res.*, 22, 5600-5607, Kayushin, A. L. *et al.*, (1996) *Nucleic Acids Res.*, 24, 3748-3755, Huse, U.S. Pat. No. 5,264,563 "PROCESS FOR SYNTHESIZING 25 OLIGONUCLEOTIDES WITH RANDOM CODONS", Lyttle *et al.*, U.S. Pat. No. 5,717,085 "PROCESS FOR PREPARING CODON AMIDITES", Shortle *et al.*, U.S. Pat. No. 5,869,644 "SYNTHESIS OF DIVERSE AND USEFUL COLLECTIONS OF 30 OLIGONUCLEOTIDES"; Greyson, U.S. Pat. No. 5,789,577 "METHOD FOR THE CONTROLLED SYNTHESIS OF POLYNUCLEOTIDE MIXTURES WHICH ENCODE DESIRED MIXTURES OF PEPTIDES"; and Huse, WO 92/06176 "SURFACE EXPRESSION LIBRARIES OF RANDOMIZED PEPTIDES".

Codon-varied oligonucleotides can be synthesized using various trinucleotide-related techniques, e.g., the trinucleotide synthesis format and the split-pool synthesis format.

Additional steps of the split-pool format optionally include splitting the extended intermediate oligonucleotide sequences into two or more separate pools. After this

Intermediate oligonucleotide sequences are produced.

step become the substrate sequences of the next repeated removing step until extended

steps are repeated, the extended oligonucleotide sequences yielded by each repeated coupling step are removed and coupling steps are optionally repeated. When these sequences, i.e. after the removal of the coupling steps, are repeated, the extended oligonucleotide sequences yielded by each repeated coupling step are repeated, the extended oligonucleotide sequences yielded by each repeated coupling step are removed and coupling steps are optionally repeated. When these

sequences. Each nucleotide has a 3' terminus, a 5' terminus, and three bases, all of which have protecting groups thereon. The coupling step yields extended oligonucleotide

substrate sequences, which are then coupled with selected trinucleotide phosphoramidite

having a δ' terminus and at least one base, both of which have protecting groups thereon.

The split-pool synthesis format includes providing substrate sequences, each

sequences, and oligonucleotides.

varied oligonucleotide is obtained. This basic synthesis format can optionally include coupling together one or more of: mononucleotides, dinucleotide phosphoramidite

steps are repeated, the extended oligonucleotide sequence yielded by each repeated coupling step becomes the substrate sequence of the next repeated removing step until a desired coding sequence is obtained.

has protecting groups thereon. The coupling step yields an extended oligonucleotide sequence. Thereafter, the removing and coupling steps are optionally repeated. When these

substate sequence, which is then coupled with a selected trimucleotide phosphoramidite sequence. The trimucleotide has a 3' terminus, a 5' terminus, and three bases, each of which

having a 3' terminus and at least one base, both of which have protecting groups thereon, having a 5' protecting group of the substrate sequence is then removed to provide a 5' deprotected

The tri-nucleotide synthesis format includes providing a substrate sequence

3. protective groups can be, e.g., β -cyanomethyl groups

group, such as *tert*-butyl dimethylsilyl (T-BDMSi), triisopropyl silyl, or trimethylsilyl. The (TMOC) or terephthaloyl, an acidic-cleavable group, such as *p*xylyl; a thionocarbonate-cleavable alkylsilyl

proteching groups such as carbonyl-containing groups, such as 9-fluorenylmethoxy carbonyl monomethoxytrityl; a carbonyl-containing group, such as carbonyl groups, such as dimethoxytrityl (DMT) or

amino phosphonamide salts-Phase chemical syntheses in which the salts of inorganic acid

oligonucleotide synthetic methods is well known to those of skill. In general, both methods

The chemistry involved in both the titricleotide and the split-pool codon-varied

is done, the 5' protecting groups of the extended intermediate oligonucleotide sequences are removed to provide 5' deprotected extended intermediate oligonucleotide sequences in the two or more separate pools. Following this, these 5' deprotected intermediates are coupled with one or more selected mononucleotides, trinucleotide phosphoramidite sequences, or 5 oligonucleotides in the two or more separate pools to yield further extended intermediate oligonucleotide sequences. In turn, these further extended sequences are pooled into a single pool. Thereafter, the steps beginning with the removal of the 5' protecting groups of the substrate sequences to provide 5' deprotected substrate sequences are optionally repeated. When these steps are repeated, the further extended oligonucleotide sequences, yielded by 10 each repeated coupling step that generates those specific sequences, become the substrate sequences of the next repeated removing step that includes those specific sequences until desired codon-varied oligonucleotides are obtained.

Both synthetic protocols described, *supra*, can optionally be performed in an automated synthesizer that automatically performs the steps. This aspect includes inputting 15 character string information into a computer, the output of which then directs the automated synthesizer to perform the steps necessary to synthesize the desired codon-varied oligonucleotides.

Further details regarding tri-nucleotide synthesis are found "USE OF CODON VARIED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch 20 et al., USSN 09/408,393, filed September 28, 1999.

TUNING NUCLEIC ACID RECOMBINATION USING OLIGONUCLEOTIDE-MEDIATED BLENDING

In one aspect, non-equimolar ratios of family shuffling oligonucleotides are used to bias recombination during the procedures noted herein. In this approach, equimolar 25 ratios of family shuffling oligonucleotides in a set of family shuffling oligonucleotides are *not* used to produce a library of recombinant nucleic acids, as in certain other methods herein. Instead, ratios of particular oligonucleotides which correspond to the sequences of a selected member or selected set of members of the family of nucleic acids from which the family shuffling oligonucleotides are derived are selected by the practitioner.

30 Thus, in one simple illustrative example, oligonucleotide mediated recombination as described herein is used to recombine, e.g., a frog gene and a human gene which are 50% identical. Family oligonucleotides are synthesized which encode both the human and the frog sequences at all polymorphic positions. However, rather than using an

acid diversity represented by the alignment. The above biasing method is optionally used to recombinant space of interest can be defined which includes all permutations of the amino structural homology (e.g., by alignment of functionally similar peptide residues). A nucleic acid level). In these cases, non-homologous sequences are "aligned" by considering similar members which display little or no detectable sequence homology (especially at the nucleic acid level). For example, the immunoglobulin superfamily includes structurally encoded proteins. For example, the immunoglobulin superfamily of structural similarity of the homologous sets of oligonucleotides derived from consideration of structural similarity of the biasing to occur. In this case, "family" oligonucleotides are substituted for non-sequence selection to proceed. In fact, sequences do not even have to be detectably homologous for of two or more related nucleic acids. Sequences do not have to be closely similar for As noted, this biasing approach can be applied to the recombination of any set method as noted, supra).

disproportionate amounts to the relevant gene synthesis method (e.g., to a PCR synthetic dispropportionate amounts of the relevant oligonucleotides, or simply supplying Biasing can be accomplished in a variety of ways, including synthesizing mediated recombination.

applicable to the recombination of any two or more nucleic acids by oligonucleotide the original frog gene. This strategy is set out in Figure 2. The strategy is generally gene, the blended gene is optionally further recombined with genes more closely similar to similarly, where the ratio of oligonucleotides is selected to produce a more frog-like blended optionally further recombined with genes more closely similar to the original human gene. oligonucleotides is selected to produce a more human-like blended gene, the blended gene is sequence to the blended gene. Thus, in the case above, in the case where the ratio of parent gene) can then be recombined with gene family members which are closely related by "blended" gene (i.e., the resulting recombinant gene with characteristics of more than one polymorphic positions can be biased to greater than 50%. In either case, the resulting wants a frog-like gene, the ratio of oligos corresponding to the frog sequence at 30%, 20%, 10%, or less of the oligos can correspond to the human sequence. Similarly, if one 90% or more of the oligos can correspond to the human sequence, with, e.g., about 40%, or polymorphic positions can be biased to greater than 50% (e.g., about 60%, 70%, 80%, or human-like gene, the ratio of oligonucleotides which correspond to the human sequence at the gene that the user wishes to emulate most closely. For example, when generating a equilibrium ratio of the human and frog derived oligonucleotides, the ratio is biased in favor of

blend the sequences with desired ratios of the nucleotides encoding relevant structurally similar amino acid sequences.

Any two or more sequences can be aligned by any algorithm or criteria of interest and the biasing method used to blend the sequences based upon any desired criteria.

5 These include sequence homology, structural similarity, predicted structural similarity (based upon any similarity criteria which are specified), or the like. It can be applied to situations in which there is a structural core that is constant, but having many structural variations built around the core (for example, an Ig domain can be a structural core having many different loop lengths and conformations being attached to the core).

10 A general advantage to this approach as compared to standard gene recombination methods is that the overall sequence identity of two sequences to be blended can be lower than the identity necessary for recombination to occur by more standard methods. In addition, sometimes only selected regions are recombined, making it possible to take any structural or functional data which is available into account in specifying how the 15 blended gene is constructed. Thus, sequence space which is not produced by some other shuffling protocols is accessed by the blended gene approach and a higher percentage of active clones can sometimes be obtained if structural information is taken into consideration. Further details regarding consideration of structural information is found in "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES AND POLYPEPTIDES 20 HAVING DESIRED CHARACTERISTICS" by Selifonov et al., attorney docket number 02-289-3US.

25 The general strategy above is applicable, e.g., to any set of genes with low sequence similarity. For example, there is a large family of TNF homologues whose sequence identity is in the range of about 30%, making standard shuffling protocols difficult to achieve. Of course, tuning recombination by selecting oligonucleotide proportions is also generally applicable to recombination of any two nucleic acids, including both high similarity homologues and low similarity homologues. Any alignment protocol can be selected to align two or more sequences and the resulting alignment can be used to create appropriate oligonucleotides to achieve recombination, and any biasing in the relative frequencies of 30 sequences as compared to parental sequences can be achieved.

TARGETS FOR OLIGONUCLEOTIDE SHUFFLING

Essentially any nucleic acid can be shuffled by the oligonucleotide mediate methods herein. No attempt is made to identify the hundreds of thousands of known nucleic

important pharmacological activities. Because of the evolutionary divergence between these appreciate that both frog and human RNases are known and are known to have a number of methods herein, particularly those methods utilizing gene blending. One of skill will 30 RNases such as Dnase and EDN are preferred targets for the synthetic

the LDL receptor ligand and corticosterone. steroid hormone receptors such as those for estrogen, progesterone, testosterone, aldosterone, transcriptional activators and suppressors, e.g., p53, Taf, Fos, Myc, Jun, Myb, Rel, and molecules and corresponding oncogenes products, e.g., Mos, Ras, Raf, and Met, and 25 CD40/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1, and hairyline/CD44, signal transduction interleukins, FGF, IGF-I, IGF-II, FGF, PDGF, TNF, TGF- α , TGF- β , EGF, KGF, SCF/c-KIT, factors, their receptors, and oncogene products, e.g., interleukins (e.g., IL-1, IL-2, IL-8, etc.), degrading RNA. Expression activators include cytokines, inflammatory molecules, growth promoters and enhancers, unwinding DNA, splicing pre-mRNA, polyadenylating RNA, and transcription factors, binding to promoters and enhancers, binding to proteins that bind to 20 transcription factors, binding to promoters and enhancers, binding to proteins that bind to binding to receptors, stimulating a signal transduction cascade, regulating expression of

expression and transcriptional activators regulate transcription by many mechanisms, e.g., by including mammals, providing a wide range of therapeutic targets. It will be appreciated that are found in prokaryotes, viruses, and eukaryotes, including fungi, plants, and animals. 15 Example transcriptional and expression activators include genes and proteins that modulate cell growth, differentiation, regulation, or the like. Expression and transcriptional activators Another class of preferred targets are transcriptional and expression activators.

catalogue and price list), and the corresponding genes are well-known. of these proteins are commercially available (See, e.g., the Sigma BioSciences 1997 oncostatin M, PD-ECF, PDGF, pleiotropin, SCF, c-kit ligand, VEGF, G-CSF etc. Many interleukins, keratinocyte growth factor, leukemia inhibitory factor, the factor, fibroblast growth factor, hepatocyte growth factor, insulin-like growth factor, the Peptide-78, GROa/MGSA, GRO β , GRO γ , MIP-1 α , MIP-1 β , MCP-1, epidemal growth hormone; growth factors and cytokines such as epithelial Neutrophil Activating 5 therapeutic proteins such as erythropoietin (EPO), insulin, peptide hormones such as human One class of preferred targets for activation includes nucleic acids encoding intermediate.

EMBL, DBJ and the NCBI. Other repositories can easily be identified by searching the acids. As noted above, common sequence repositories for known proteins include GenBank

genes, oligonucleotide-mediated recombination methods are particularly useful in recombining the nucleic acids.

Similarly, proteins from infectious organisms for possible vaccine applications, described in more detail below, including infectious fungi, e.g., *Aspergillus*, 5 *Candida* species; bacteria, particularly *E. coli*, which serves a model for pathogenic bacteria, as well as medically important bacteria such as *Staphylococci* (e.g., *aureus*), *Streptococci* (e.g., *pneumoniae*), *Clostridia* (e.g., *perfringens*), *Neisseria* (e.g., *gonorrhoea*), *Enterobacteriaceae* (e.g., *coli*), *Helicobacter* (e.g., *pylori*), *Vibrio* (e.g., *cholerae*), *Campylobacter* (e.g., *jejuni*), *Pseudomonas* (e.g., *aeruginosa*), *Haemophilus* (e.g., 10 *influenzae*), *Bordetella* (e.g., *pertussis*), *Mycoplasma* (e.g., *pneumoniae*), *Ureaplasma* (e.g., *urealyticum*), *Legionella* (e.g., *pneumophila*), *Spirochetes* (e.g., *Treponema*, *Leptospira*, and *Borrelia*), *Mycobacteria* (e.g., *tuberculosis*, *smegmatis*), *Actinomyces* (e.g., *israelii*), *Nocardia* (e.g., *asteroides*), *Chlamydia* (e.g., *trachomatis*), *Rickettsia*, *Coxiella*, *Ehrlichia*, *Rocholimaea*, *Brucella*, *Yersinia*, *Francisella*, and *Pasteurella*; protozoa such as sporozoa 15 (e.g., *Plasmodia*), rhizopods (e.g., *Entamoeba*) and flagellates (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); viruses such as (+) RNA viruses (examples include Poxviruses e.g., *vaccinia*; Picornaviruses, e.g. *polio*; Togaviruses, e.g., *rubella*; Flaviviruses, e.g., HCV; and Coronaviruses), (-) RNA viruses (examples include Rhabdoviruses, e.g., VSV; Paramyxoviruses, e.g., RSV; Orthomyxoviruses, e.g., influenza; Bunyaviruses; and 20 Arenaviruses), dsDNA viruses (Reoviruses, for example), RNA to DNA viruses, i.e., Retroviruses, e.g., especially HIV and HTLV, and certain DNA to RNA viruses such as Hepatitis B virus.

Other proteins relevant to non-medical uses, such as inhibitors of transcription or toxins of crop pests e.g., insects, fungi, weed plants, and the like, are also preferred targets 25 for oligonucleotide shuffling. Industrially important enzymes such as monooxygenases (e.g., p450s), proteases, nucleases, and lipases are also preferred targets. As an example, subtilisin can be evolved by shuffling family oligonucleotides for homologous forms of the gene for subtilisin. Von der Osten et al., *J. Biotechnol.* 28:55-68 (1993) provide an example subtilisin coding nucleic acids and additional nucleic acids are present in GENBANK®. Proteins 30 which aid in folding such as the chaperonins are also preferred targets.

Preferred known genes suitable for oligonucleotide mediated shuffling also include the following: Alpha-1 antitrypsin, Angiostatin, Antihemolytic factor, Apolipoprotein, Apoprotein, Atrial natriuretic factor, Atrial natriuretic polypeptide, Atrial peptides, C-X-C chemokines (e.g., T39765, NAP-2, ENA-78, Gro-a, Gro-b, Gro-c, IP-10,

GC-2, NAP-4, SDF-1, PFA, MIG), Calcium, CC chemokines (e.g., Monocyte chemoattractant protein-1, Monocyte chemoattractant protein-2, Monocyte chemoattractant protein-3, Monocyte inflammatory protein-1 alpha, Monocyte inflammatory protein-1 beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262), CD40 ligand, Collagen, Colony stimulating factor (CSF), Complement factor 5a, Complement inhibitor, Complement receptor 1, Factor IX, Factor VII, Factor VIII, Factor X, Fibronectin, Fibronectin, Glucoceerebrosidase, Gonadotropin, Hedgehog proteins (e.g., Sonic, Indian, Desert), Hemoglobin (for blood substitute; for radiosensitization), Hirudin, Human serum albumin, Lactoferrin, Luciferase, Neutrophil, Neutrophil inhibitory factor (NIF), Osteogenic factor, Parathyroid hormone, Protein A, Protein G, Relaxin, Renin, Salmon calcitonin, Interleukin receptors (IL-1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15), Soluble TNF receptor, Somatotropin, Somatostatin, Somatotropin, Strepotokinase, Superantigens, i.e., Staphylococcal enterotoxins (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), Toxic shock syndrome toxin (TSST-1), Exfoliating toxins A and B, Pyrogenic exotoxins A, B, and C, and M, arthritides mitogen, Superoxide dismutase, Thyrosin alpha 1, Tissue plasminogen activator, Tumor necrosis factor beta (TNF beta), Tumor necrosis factor receptor (TNFR), encoding small proteins are particularly preferred targets, because conventional recombination methods provide only limited product sequence diversity. This is because sequencing small proteins such as defensins (antifungal proteins of about 50 amino acids, EFA0 (an anti fungal protein of 28 amino acids), peptide antibiotics, and peptide insecticidal proteins are also preferred targets and exist as families of related proteins. Nucleic acids combining small proteins are particularly preferred targets, because conventional recombination methods provide only limited product sequence diversity. This is because sequencing small proteins such as defensins (antifungal proteins of about 50 amino acids, 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995 1000 1005 1010 1015 1020 1025 1030 1035 1040 1045 1050 1055 1060 1065 1070 1075 1080 1085 1090 1095 1100 1105 1110 1115 1120 1125 1130 1135 1140 1145 1150 1155 1160 1165 1170 1175 1180 1185 1190 1195 1200 1205 1210 1215 1220 1225 1230 1235 1240 1245 1250 1255 1260 1265 1270 1275 1280 1285 1290 1295 1300 1305 1310 1315 1320 1325 1330 1335 1340 1345 1350 1355 1360 1365 1370 1375 1380 1385 1390 1395 1400 1405 1410 1415 1420 1425 1430 1435 1440 1445 1450 1455 1460 1465 1470 1475 1480 1485 1490 1495 1500 1505 1510 1515 1520 1525 1530 1535 1540 1545 1550 1555 1560 1565 1570 1575 1580 1585 1590 1595 1600 1605 1610 1615 1620 1625 1630 1635 1640 1645 1650 1655 1660 1665 1670 1675 1680 1685 1690 1695 1700 1705 1710 1715 1720 1725 1730 1735 1740 1745 1750 1755 1760 1765 1770 1775 1780 1785 1790 1795 1800 1805 1810 1815 1820 1825 1830 1835 1840 1845 1850 1855 1860 1865 1870 1875 1880 1885 1890 1895 1900 1905 1910 1915 1920 1925 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020 2025 2030 2035 2040 2045 2050 2055 2060 2065 2070 2075 2080 2085 2090 2095 2100 2105 2110 2115 2120 2125 2130 2135 2140 2145 2150 2155 2160 2165 2170 2175 2180 2185 2190 2195 2200 2205 2210 2215 2220 2225 2230 2235 2240 2245 2250 2255 2260 2265 2270 2275 2280 2285 2290 2295 2300 2305 2310 2315 2320 2325 2330 2335 2340 2345 2350 2355 2360 2365 2370 2375 2380 2385 2390 2395 2400 2405 2410 2415 2420 2425 2430 2435 2440 2445 2450 2455 2460 2465 2470 2475 2480 2485 2490 2495 2500 2505 2510 2515 2520 2525 2530 2535 2540 2545 2550 2555 2560 2565 2570 2575 2580 2585 2590 2595 2600 2605 2610 2615 2620 2625 2630 2635 2640 2645 2650 2655 2660 2665 2670 2675 2680 2685 2690 2695 2700 2705 2710 2715 2720 2725 2730 2735 2740 2745 2750 2755 2760 2765 2770 2775 2780 2785 2790 2795 2800 2805 2810 2815 2820 2825 2830 2835 2840 2845 2850 2855 2860 2865 2870 2875 2880 2885 2890 2895 2900 2905 2910 2915 2920 2925 2930 2935 2940 2945 2950 2955 2960 2965 2970 2975 2980 2985 2990 2995 3000 3005 3010 3015 3020 3025 3030 3035 3040 3045 3050 3055 3060 3065 3070 3075 3080 3085 3090 3095 3100 3105 3110 3115 3120 3125 3130 3135 3140 3145 3150 3155 3160 3165 3170 3175 3180 3185 3190 3195 3200 3205 3210 3215 3220 3225 3230 3235 3240 3245 3250 3255 3260 3265 3270 3275 3280 3285 3290 3295 3300 3305 3310 3315 3320 3325 3330 3335 3340 3345 3350 3355 3360 3365 3370 3375 3380 3385 3390 3395 3400 3405 3410 3415 3420 3425 3430 3435 3440 3445 3450 3455 3460 3465 3470 3475 3480 3485 3490 3495 3500 3505 3510 3515 3520 3525 3530 3535 3540 3545 3550 3555 3560 3565 3570 3575 3580 3585 3590 3595 3600 3605 3610 3615 3620 3625 3630 3635 3640 3645 3650 3655 3660 3665 3670 3675 3680 3685 3690 3695 3700 3705 3710 3715 3720 3725 3730 3735 3740 3745 3750 3755 3760 3765 3770 3775 3780 3785 3790 3795 3800 3805 3810 3815 3820 3825 3830 3835 3840 3845 3850 3855 3860 3865 3870 3875 3880 3885 3890 3895 3900 3905 3910 3915 3920 3925 3930 3935 3940 3945 3950 3955 3960 3965 3970 3975 3980 3985 3990 3995 4000 4005 4010 4015 4020 4025 4030 4035 4040 4045 4050 4055 4060 4065 4070 4075 4080 4085 4090 4095 4100 4105 4110 4115 4120 4125 4130 4135 4140 4145 4150 4155 4160 4165 4170 4175 4180 4185 4190 4195 4200 4205 4210 4215 4220 4225 4230 4235 4240 4245 4250 4255 4260 4265 4270 4275 4280 4285 4290 4295 4300 4305 4310 4315 4320 4325 4330 4335 4340 4345 4350 4355 4360 4365 4370 4375 4380 4385 4390 4395 4400 4405 4410 4415 4420 4425 4430 4435 4440 4445 4450 4455 4460 4465 4470 4475 4480 4485 4490 4495 4500 4505 4510 4515 4520 4525 4530 4535 4540 4545 4550 4555 4560 4565 4570 4575 4580 4585 4590 4595 4600 4605 4610 4615 4620 4625 4630 4635 4640 4645 4650 4655 4660 4665 4670 4675 4680 4685 4690 4695 4700 4705 4710 4715 4720 4725 4730 4735 4740 4745 4750 4755 4760 4765 4770 4775 4780 4785 4790 4795 4800 4805 4810 4815 4820 4825 4830 4835 4840 4845 4850 4855 4860 4865 4870 4875 4880 4885 4890 4895 4900 4905 4910 4915 4920 4925 4930 4935 4940 4945 4950 4955 4960 4965 4970 4975 4980 4985 4990 4995 5000 5005 5010 5015 5020 5025 5030 5035 5040 5045 5050 5055 5060 5065 5070 5075 5080 5085 5090 5095 5100 5105 5110 5115 5120 5125 5130 5135 5140 5145 5150 5155 5160 5165 5170 5175 5180 5185 5190 5195 5200 5205 5210 5215 5220 5225 5230 5235 5240 5245 5250 5255 5260 5265 5270 5275 5280 5285 5290 5295 5300 5305 5310 5315 5320 5325 5330 5335 5340 5345 5350 5355 5360 5365 5370 5375 5380 5385 5390 5395 5400 5405 5410 5415 5420 5425 5430 5435 5440 5445 5450 5455 5460 5465 5470 5475 5480 5485 5490 5495 5500 5505 5510 5515 5520 5525 5530 5535 5540 5545 5550 5555 5560 5565 5570 5575 5580 5585 5590 5595 5600 5605 5610 5615 5620 5625 5630 5635 5640 5645 5650 5655 5660 5665 5670 5675 5680 5685 5690 5695 5700 5705 5710 5715 5720 5725 5730 5735 5740 5745 5750 5755 5760 5765 5770 5775 5780 5785 5790 5795 5800 5805 5810 5815 5820 5825 5830 5835 5840 5845 5850 5855 5860 5865 5870 5875 5880 5885 5890 5895 5900 5905 5910 5915 5920 5925 5930 5935 5940 5945 5950 5955 5960 5965 5970 5975 5980 5985 5990 5995 6000 6005 6010 6015 6020 6025 6030 6035 6040 6045 6050 6055 6060 6065 6070 6075 6080 6085 6090 6095 6100 6105 6110 6115 6120 6125 6130 6135 6140 6145 6150 6155 6160 6165 6170 6175 6180 6185 6190 6195 6200 6205 6210 6215 6220 6225 6230 6235 6240 6245 6250 6255 6260 6265 6270 6275 6280 6285 6290 6295 6300 6305 6310 6315 6320 6325 6330 6335 6340 6345 6350 6355 6360 6365 6370 6375 6380 6385 6390 6395 6400 6405 6410 6415 6420 6425 6430 6435 6440 6445 6450 6455 6460 6465 6470 6475 6480 6485 6490 6495 6500 6505 6510 6515 6520 6525 6530 6535 6540 6545 6550 6555 6560 6565 6570 6575 6580 6585 6590 6595 6600 6605 6610 6615 6620 6625 6630 6635 6640 6645 6650 6655 6660 6665 6670 6675 6680 6685 6690 6695 6700 6705 6710 6715 6720 6725 6730 6735 6740 6745 6750 6755 6760 6765 6770 6775 6780 6785 6790 6795 6800 6805 6810 6815 6820 6825 6830 6835 6840 6845 6850 6855 6860 6865 6870 6875 6880 6885 6890 6895 6900 6905 6910 6915 6920 6925 6930 6935 6940 6945 6950 6955 6960 6965 6970 6975 6980 6985 6990 6995 7000 7005 7010 7015 7020 7025 7030 7035 7040 7045 7050 7055 7060 7065 7070 7075 7080 7085 7090 7095 7100 7105 7110 7115 7120 7125 7130 7135 7140 7145 7150 7155 7160 7165 7170 7175 7180 7185 7190 7195 7200 7205 7210 7215 7220 7225 7230 7235 7240 7245 7250 7255 7260 7265 7270 7275 7280 7285 7290 7295 7300 7305 7310 7315 7320 7325 7330 7335 7340 7345 7350 7355 7360 7365 7370 7375 7380 7385 7390 7395 7400 7405 7410 7415 7420 7425 7430 7435 7440 7445 7450 7455 7460 7465 7470 7475 7480 7485 7490 7495 7500 7505 7510 7515 7520 7525 7530 7535 7540 7545 7550 7555 7560 7565 7570 7575 7580 7585 7590 7595 7600 7605 7610 7615 7620 7625 7630 7635 7640 7645 7650 7655 7660 7665 7670 7675 7680 7685 7690 7695 7700 7705 7710 7715 7720 7725 7730 7735 7740 7745 7750 7755 7760 7765 7770 7775 7780 7785 7790 7795 7800 7805 7810 7815 7820 7825 7830 7835 7840 7845 7850 7855 7860 7865 7870 7875 7880 7885 7890 7895 7900 7905 7910 7915 7920 7925 7930 7935 7940 7945 7950 7955 7960 7965 7970 7975 7980 7985 7990 7995 8000 8005 8010 8015 8020 8025 8030 8035 8040 8045 8050 8055 8060 8065 8070 8075 8080 8085 8090 8095 8100 8105 8110 8115 8120 8125 8130 8135 8140 8145 8150 8155 8160 8165 8170 8175 8180 8185 8190 8195 8200 8205 8210 8215 8220 8225 8230 8235 8240 8245 8250 8255 8260 8265 8270 8275 8280 8285 8290 8295 8300 8305 8310 8315 8320 8325 8330 8335 8340 8345 8350 8355 8360 8365 8370 8375 8380 8385 8390 8395 8400 8405 8410 8415 8420 8425 8430 8435 8440 8445 8450 8455 8460 8465 8470 8475 8480 8485 8490 8495 8500 8505 8510 8515 8520 8525 8530 8535 8540 8545 8550 8555 8560 8565 8570 8575 8580 8585 8590 8595 8600 8605 8610 8615 8620 8625 8630 8635 8640 8645 8650 8655 8660 8665 8670 8675 8680 8685 8690 8695 8700 8705 8710 8715 8720 8725 8730 8735 8740 8745 8750 8755 8760 8765 8770 8775 8780 8785 8790 8795 8800 8805 8810 8815 8820 8825 8830 8835 8840 8845 8850 8855 8860 8865 8870 8875 8880 8885 8890 8895 8900 8905 8910 8915 8920 8925 8930 8935 8940 8945 8950 8955 8960 8965 8970 8975 8980 8985 8990 8995 9000 9005 9010 9015 9020 9025 9030 9035 9040 9045 9050 9055 9060 9065 9070 9075 9080 9085 9090 9095 9100 9105 9110 9115 9120 9125 9130 9135 9140 9145 9150 9155 9160 9165 9170 9175 9180 9185 9190 9195 9200 9205 9210 9215 9220 9225 9230 9235 9240 9245 9250 9255 9260 9265 9270 9275 9280 9285 9290 9295 9300 9305 9310 9315 9320 9325 9330 9335 9340 9345 9350 9355 9360 9365 9370 9375 9380 9385 9390 9395 9400 9405 9410 9415 9420 9425 9430 9435 9440 9445 9450 9455 9460 9465 9470 9475 9480 9485 9490 9495 9500 9505 9510 9515 9520 9525 9530 9535 9540 9545 9550 9555 9560 9565 9570 9575 9580 9585 9590 9595 9600 9605 9610 9615 9620 9625 9630 9635 9640 9645 9650 9655 9660 9665 9670 9675 9680 9685 9690 9695 9700 9705 9710 9715 9720 9725 9730 9735 9740 9745 9750 9755 9760 9765 9770 9775 9780 9785 9790 9795 9800 9805 9810 9815 9820 9825 9830 9835 9840 9845 9850 9855 9860 9865 9870 9875 9880 9885 9890 9895 9900 9905 9910 9915 9920 9925 9930 9935 9940 9945 9950 9955 9960 9965 9970 9975 9980 9985 9990 9995 9999

DNA SHUFFLING AND GENE REASSEMBLY—HYBRID SYNTHETIC SHUFFLING METHODS

One aspect of the present invention is the ability to use family shuffling oligonucleotides and cross over oligonucleotides as recombination templates/intermediates in various DNA shuffling methods. In addition, nucleic acids made by the new synthetic techniques herein can be reshuffled by other available shuffling methodologies.

A variety of such methods are known, including those taught by the inventors and their coworkers. The following publications describe a variety of recursive recombination procedures and/or related methods which can be practiced in conjunction with the processes of the invention: Stemmer, et al., (1999) "Molecular breeding of viruses for targeting and other clinical properties. Tumor Targeting" 4:1-4; Nasset al. (1999) "DNA Shuffling of subgenomic sequences of subtilisin" Nature Biotechnology 17:893-896; Chang et al. (1999) "Evolution of a cytokine using DNA family shuffling" Nature Biotechnology 17:793-797; Minshull and Stemmer (1999) "Protein evolution by molecular breeding" Current Opinion in Chemical Biology 3:284-290; Christians et al. (1999) "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling" Nature Biotechnology 17:259-264; Crameriet al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" Nature 391:288-291; Crameri et al. (1997) "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," Nature Biotechnology 15:436-438; Zhang et al. (1997) "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening" Proceedings of the National Academy of Sciences, U.S.A. 94:4504-4509; Patten et al. (1997) "Applications of DNA Shuffling to Pharmaceuticals and Vaccines" Current Opinion in Biotechnology 8:724-733; Crameri et al. (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" Nature Medicine 2:100-103; Crameri et al. (1996) "Improved green fluorescent protein by molecular evolution using DNA shuffling" Nature Biotechnology 14:315-319; Gates et al. (1996) "Affinity selective isolation of ligands from peptide libraries through display on a lac repressor 'headpiece dimer'" Journal of Molecular Biology 255:373-386; Stemmer (1996) "Sexual PCR and Assembly PCR" In: The Encyclopedia of Molecular Biology. VCH Publishers, New York. pp.447-457; Crameri and Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes" BioTechniques 18:194-195; Stemmer et al.; (1995) "Single-step assembly of a gene and entire plasmid form large numbers of oligodeoxyribonucleotides" Gene, 164:49-53; Stemmer (1995) "The Evolution of Molecular Computation" Science 270: 1510; Stemmer

Additional details regarding DNA shuffling methods are found in U.S. Patents by the inventors and their co-workers, including: United States Patent 5,605,793 to Stemberger (February 25, 1997), "METHODS FOR IN VITRO RECOMBINANTION"; United States Patent 5,811,238 to Stemberger et al. (September 22, 1998) "METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESERED CHARACTERISTICS BY ITERATIVE REACTIONS"; United States Patent 5,834,252 to Stemberger, et al. (November 10, 1998) "END-COMPLEMENTARY POLYMERASE REACTION"; and United States Patent 5,837,458 to Minschull, et al. (November 17, 1998), "METHODS AND COMPOSITIONS FOR CELLULAR AND METABOLIC ENGINEERING". In addition, details and formats for nucleic acid shuffling are found in a variety of PCT and foreign patent applications, including: Stemberger and Crameri, "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY" WO 95/22625; Stemberger and Lipschutz, "END COMPLEMENTARY POLYMERASE CHAIN REACTION" WO 96/33207; Stemberger and Crameri, "METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESERED CHARACTERISTICS BY ITERATIVE REACTIONS" WO 97/0078; Minschull and Stemberger, "METHODS AND COMPOSITIONS FOR CELLULAR AND METABOLIC ENGINEERING" WO 97/35966; Pumponen et al., "TARGETING OF GENETIC VACCINE VECTORS" WO 99/41402; Pumponen et al., "ANTIGEN LIBRARY IMMUNIZATION" WO 99/41383; Pumponen et al., "GENETIC VACCINE VECTOR ENGINEERING" WO 99/41369; Pumponen et al., "OPTIMIZATION OF IMMUNOMODULATOR PROPERTIES OF GENETIC VACCINES" WO 9941368; Stemberger and Crameri, "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY" EP 0934999; Stemberger et al., "EVOLVING CELLULAR DNA UPTAKE BY RECURSIVE SEQUENCING" EP 0932670; Stemberger et al., "MODIFICATION OF VIRUS RECOMBINANTION" EP 0932670; Stemberger et al., "HOST RANGE BY VIRTUAL GENOME SHUFFLING" WO 9923107; and et al., "HUMAN PAPILLOMAVIRUS VECTORS" WO 9921979; Del Cardayre et al.

“EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION” WO 9831837; Patten and Stemmer, “METHODS AND COMPOSITIONS FOR POLYPEPTIDE ENGINEERING” WO 9827230; Stemmer et al., and “METHODS FOR OPTIMIZATION OF GENE THERAPY BY RECURSIVE SEQUENCE SHUFFLING AND SELECTION” WO9813487.

5 Certain U.S. Applications provide additional details regarding DNA shuffling and related techniques, including “SHUFFLING OF CODON ALTERED GENES” by Patten et al. filed September 29, 1998, (USSN 60/102,362), January 29, 1999 (USSN 60/117,729), and September 28, 1999, USSN PCT/US99/22588; “EVOLUTION OF WHOLE CELLS 10 AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION”, by del Cardyre et al. filed July 15, 1998 (USSN 09/166,188), and July 15, 1999 (USSN 09/354,922); “OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION” by Crameri et al., filed February 5, 1999 (USSN 60/118,813) and filed June 24, 1999 (USSN 60/141,049) and filed September 28, 1999 (USSN 09/408,392), and “USE OF CODON-BASED 15 OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING” by Welch et al., filed September 28, 1999 (USSN 09/408,393). Finally, the applications cited above in the section entitled “Cross Reference to Related Applications” provide relevant formats.

The foregoing references also provide additional details on the process of hybridizing and elongating nucleic acids to achieve nucleic acid recombination.

20 In one aspect, a hybrid method which uses family gene shuffling in combination with more traditional recombination based shuffling methods is used. For example, an active nucleic acid can be reassembled from oligonucleotides to have a few or even no homologous substitutions relative to a given target gene. The reassembled “backbone” nucleic acid is treated with DNase as in standard methods, and the resulting 25 DNased fragments are spiked with family oligonucleotides comprising sequences corresponding to regions of sequence identity and diversity in a given nucleic acid. The nucleic acids are then reassembled into a library of homologous sequences by the methods below (e.g., PCR reassembly, or other reassembly methods). This procedure can result in an increase in the percentage of active clones which are found as compared to oligonucleotides 30 synthetic methods which do not incorporate the use of a backbone nucleic acid.

A number of the publications of the inventors and their co-workers, as well as other investigators in the art describe techniques which facilitate DNA shuffling, e.g., by providing for reassembly of genes from small fragments, including oligonucleotides, as relevant to the present invention. For example, Stemmer et al. (1998) U.S. Pat. No.

5,834,252 END COMPLEMENTARY POLYMERASE REACTION describes processes for amplifying and detecting a target sequence (e.g., in a mixture of nucleic acids), as well as for assembling large polymeric fragments. Crameri et al. (1998) *Nature* 391: 288-291 provides basic methodologies for gene reassembly, as does Crameri et al. (1998) *Bio techniques* 18(2): 194-196.

Other diversity generating approaches can also be used to modify nucleic acids produced by the methods herein, or to be used as templates for the methods herein. For example, additional diversity can be introduced by methods which result in the alteration of individual nucleotides or groups of contiguous or non-contiguous nucleotides, i.e., oligonucleotide directed mutagenesis (Zoller and Smith, *Nucl. Acids Res.* 10: 6487-6500 (1982), *Methods in Enzymol.* 100: 468-500 (1983), and *Methods in Enzymol.* 154: 329-350 (1982), *Methods in Enzymol.* 100: 468-500 (1983), and *Methods in Enzymol.* 154: 367 (1987); Krammer et al., *Nucl. Acids Res.* 16: 6987-6999 (1988)). Additional methods include point mismatch repair (Krammer et al., *Cell* 38: 879-887 (1984)), mutagenesis using repair-deficient host strains (Krammer et al., *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, *Methods in Enzymol.* 154: 382-403 (1987)), deletion mutagenesis (Eghbaderzadeh and Hemikoff, *Nucl. Acids Res.* 14: 5115 (1986)), restriction selection and restriction-purification (Wells et al., *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986)), mutagenesis by total gene synthesis (Nambiar et al., *Science* 223: 1299-1301 (1984); Sakkam and Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells et al., *Gene* 34: 315-323 (1985); and Grunstrom et al., *Nucl. Acids Res.* 13: 3305-3316 (1986)).

30 (Carter et al., *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, *Methods in Enzymol.* 154: 382-403 (1987)), deletion mutagenesis (Eghbaderzadeh and Hemikoff, *Nucl. Acids Res.* 14: 5115 (1986)), restriction selection and restriction-purification (Wells et al., *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986)), mutagenesis by total gene synthesis (Nambiar et al., *Science* 223: 1299-1301 (1984); Sakkam and Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells et al., *Gene* 34: 315-323 (1985); and Grunstrom et al., *Nucl. Acids Res.* 13: 3305-3316 (1986)).

25 (Krammer et al., *Cell* 38: 879-887 (1984)), mutagenesis using repair-deficient host strains (Krammer et al., *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, *Methods in Enzymol.* 154: 382-403 (1987)), deletion mutagenesis (Eghbaderzadeh and Hemikoff, *Nucl. Acids Res.* 14: 5115 (1986)), restriction selection and restriction-purification (Wells et al., *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986)), mutagenesis by total gene synthesis (Nambiar et al., *Science* 223: 1299-1301 (1984); Sakkam and Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells et al., *Gene* 34: 315-323 (1985); and Grunstrom et al., *Nucl. Acids Res.* 13: 3305-3316 (1986)).

20 (Eckstein, *Nucl. Acids Res.* 14: 9679-9698 (1986); Sayers et al., *Nucl. Acids Res.* 16: 791-802 8749-8764 (1985); Taylor et al., *Nucl. Acids Res.* 13: 8765-8787 (1985); Nakamaye and (1987) phosphothioate-modified DNA mutagenesis (Taylor et al., *Nucl. Acids Res.* 13: 8765-8787 (1985); Sayers et al., *Nucl. Acids Res.* 16: 803-814 (1988)), mutagenesis using uracil- containing templates (Kunkel, *Proc. Natl. Acad. Sci. USA* 82: 488-492 (1985) and Kunkel et al., *Nucl. Acids Res.* 12: 9441-9456 (1984); Krammer and Fritz, *Methods in Enzymol.* 154: 367 (1987); Krammer et al., *Nucl. Acids Res.* 16: 7207 (1988)); and Fritz et al., *Nucl. Acids Res.* 16: 6987-6999 (1988)). Additional methods include point mismatch repair (Krammer et al., *Cell* 38: 879-887 (1984)), mutagenesis using repair-deficient host strains (Krammer et al., *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, *Methods in Enzymol.* 154: 382-403 (1986)), deletion mutagenesis (Eghbaderzadeh and Hemikoff, *Nucl. Acids Res.* 14: 5115 (1986)), restriction selection and restriction-purification (Wells et al., *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986)), mutagenesis by total gene synthesis (Nambiar et al., *Science* 223: 1299-1301 (1984); Sakkam and Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells et al., *Gene* 34: 315-323 (1985); and Grunstrom et al., *Nucl. Acids Res.* 13: 3305-3316 (1986)).

15 (Lilley, eds., Springer Verlag, Berlin (1987)). Included among these methods are oligonucleotide-directed mutagenesis (Zoller and Smith, *Nucl. Acids Res.* 10: 6487-6500 (1982), *Methods in Enzymol.* 100: 468-500 (1983), and *Methods in Enzymol.* 154: 329-350 (1982), *Methods in Enzymol.* 100: 468-500 (1983), and *Methods in Enzymol.* 154: 367 (1987); Krammer et al., *Nucl. Acids Res.* 12: 9441-9456 (1984); Krammer and Fritz, *Methods in Enzymol.* 154: 367 (1987); Krammer et al., *Nucl. Acids Res.* 16: 7207 (1988)); and Kunkel, *Proc. Natl. Acad. Sci. USA* 82: 488-492 (1985) and Kunkel et al., *Nucl. Acids Res.* 12: 9441-9456 (1984); Krammer and Fritz, *Methods in Enzymol.* 154: 367 (1987); Krammer et al., *Nucl. Acids Res.* 16: 7207 (1988)); and Fritz et al., *Nucl. Acids Res.* 16: 6987-6999 (1988)). Additional methods include point mismatch repair (Krammer et al., *Cell* 38: 879-887 (1984)), mutagenesis using repair-deficient host strains (Krammer et al., *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, *Methods in Enzymol.* 154: 382-403 (1986)), deletion mutagenesis (Eghbaderzadeh and Hemikoff, *Nucl. Acids Res.* 14: 5115 (1986)), restriction selection and restriction-purification (Wells et al., *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986)), mutagenesis by total gene synthesis (Nambiar et al., *Science* 223: 1299-1301 (1984); Sakkam and Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells et al., *Gene* 34: 315-323 (1985); and Grunstrom et al., *Nucl. Acids Res.* 13: 3305-3316 (1986)).

10 (PCT/US98/05223; Publ. No. WO98/42727); oligonucleotide-directed mutagenesis (for review see, Smith, Ann. Rev. Genet. 19: 423-462 (1985); Boistein and Shorlie, *Science* 229: 1193-1201 (1985); Carter, *Biochem. J.* 237: 1-7 (1986); Kunkel, "The efficiency of oligonucleotide directed mutagenesis" in *Nucleic acids & Molecular Biology*, Eckstein and Lilley, eds., Springer Verlag, Berlin (1987)). Included among these methods are oligonucleotide directed mutagenesis (Zoller and Smith, *Nucl. Acids Res.* 10: 6487-6500 (1982), *Methods in Enzymol.* 100: 468-500 (1983), and *Methods in Enzymol.* 154: 329-350 (1982), *Methods in Enzymol.* 100: 468-500 (1983), and *Methods in Enzymol.* 154: 367 (1987); Krammer et al., *Nucl. Acids Res.* 12: 9441-9456 (1984); Krammer and Fritz, *Methods in Enzymol.* 154: 367 (1987); Krammer et al., *Nucl. Acids Res.* 16: 7207 (1988)); and Kunkel, *Proc. Natl. Acad. Sci. USA* 82: 488-492 (1985) and Kunkel et al., *Nucl. Acids Res.* 12: 9441-9456 (1984); Krammer and Fritz, *Methods in Enzymol.* 154: 367 (1987); Krammer et al., *Nucl. Acids Res.* 16: 7207 (1988)); and Fritz et al., *Nucl. Acids Res.* 16: 6987-6999 (1988)). Additional methods include point mismatch repair (Krammer et al., *Cell* 38: 879-887 (1984)), mutagenesis using repair-deficient host strains (Krammer et al., *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, *Methods in Enzymol.* 154: 382-403 (1986)), deletion mutagenesis (Eghbaderzadeh and Hemikoff, *Nucl. Acids Res.* 14: 5115 (1986)), restriction selection and restriction-purification (Wells et al., *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986)), mutagenesis by total gene synthesis (Nambiar et al., *Science* 223: 1299-1301 (1984); Sakkam and Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells et al., *Gene* 34: 315-323 (1985); and Grunstrom et al., *Nucl. Acids Res.* 13: 3305-3316 (1986)).

5 Other diversity generating approaches can also be used to modify nucleic acids produced by the methods herein, or to be used as templates for the methods herein. For example, additional diversity can be introduced by methods which result in the alteration of individual nucleotides or groups of contiguous or non-contiguous nucleotides, i.e., oligonucleotide directed mutagenesis. Mutagenesis methods include, for example, recombination techniques or groups of contiguous or non-contiguous nucleotides, i.e., individual nucleotides or groups of contiguous or non-contiguous nucleotides, i.e., (PCT/US98/05223; Publ. No. WO98/42727); oligonucleotide-directed mutagenesis (for review see, Smith, Ann. Rev. Genet. 19: 423-462 (1985); Boistein and Shorlie, *Science* 229: 1193-1201 (1985); Carter, *Biochem. J.* 237: 1-7 (1986); Kunkel, "The efficiency of oligonucleotide directed mutagenesis" in *Nucleic acids & Molecular Biology*, Eckstein and Lilley, eds., Springer Verlag, Berlin (1987)). Included among these methods are oligonucleotide directed mutagenesis (Zoller and Smith, *Nucl. Acids Res.* 10: 6487-6500 (1982), *Methods in Enzymol.* 100: 468-500 (1983), and *Methods in Enzymol.* 154: 329-350 (1982), *Methods in Enzymol.* 100: 468-500 (1983), and *Methods in Enzymol.* 154: 367 (1987); Krammer et al., *Nucl. Acids Res.* 12: 9441-9456 (1984); Krammer and Fritz, *Methods in Enzymol.* 154: 367 (1987); Krammer et al., *Nucl. Acids Res.* 16: 7207 (1988)); and Kunkel, *Proc. Natl. Acad. Sci. USA* 82: 488-492 (1985) and Kunkel et al., *Nucl. Acids Res.* 12: 9441-9456 (1984); Krammer and Fritz, *Methods in Enzymol.* 154: 367 (1987); Krammer et al., *Nucl. Acids Res.* 16: 7207 (1988)); and Fritz et al., *Nucl. Acids Res.* 16: 6987-6999 (1988)). Additional methods include point mismatch repair (Krammer et al., *Cell* 38: 879-887 (1984)), mutagenesis using repair-deficient host strains (Krammer et al., *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, *Methods in Enzymol.* 154: 382-403 (1986)), deletion mutagenesis (Eghbaderzadeh and Hemikoff, *Nucl. Acids Res.* 14: 5115 (1986)), restriction selection and restriction-purification (Wells et al., *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986)), mutagenesis by total gene synthesis (Nambiar et al., *Science* 223: 1299-1301 (1984); Sakkam and Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells et al., *Gene* 34: 315-323 (1985); and Grunstrom et al., *Nucl. Acids Res.* 13: 3305-3316 (1986)).

(1985). Kits for mutagenesis are commercially available (e.g., Bio-Rad, Amersham International, Anglian Biotechnology).

Other diversity generation procedures are proposed in U.S. Patent No. 5,756,316; U.S. Patent No. 5,965,408; Ostermeier et al. (1999) "A combinatorial approach to hybrid enzymes independent of DNA homology" *Nature Biotech* 17:1205; U.S. Patent No. 5,783,431; U.S. Patent No. 5,824,485; U.S. Patent 5,958,672; Jirholt et al. (1998) "Exploiting sequence space: shuffling in vivo formed complementarity determining regions into a master framework" *Gene* 215: 471; U.S. Patent No. 5,939,250; WO 99/10539; WO 98/58085; WO 99/10539 and others. These diversity generating methods can be combined with each other or with shuffling reactions or oligo shuffling methods, in any combination selected by the user, to produce nucleic acid diversity, which may be screened for using any available screening method.

Following recombination or other diversification reactions, any nucleic acids which are produced can be selected for a desired activity. In the context of the present invention, this can include testing for and identifying any detectable or assayable activity, by any relevant assay in the art. A variety of related (or even unrelated) properties can be assayed for, using any available assay.

DNA SHUFFLING WITHOUT THE USE OF PCR

Although one preferred format for gene reassembly uses PCR, other formats are also useful. For example, site-directed or oligonucleotide-directed mutagenesis methods can be used to generate chimeras between 2 or more parental genes (whether homologous or non-homologous). In this regard, one aspect of the present invention relates to a new method of performing recombination between nucleic acids by ligation of libraries of oligonucleotides corresponding to the nucleic acids to be recombined.

In this format, a set of a plurality of oligonucleotides which includes a plurality of nucleic acid sequences from a plurality of the parental nucleic acids are ligated to produce one or more recombinant nucleic acid(s), typically encoding a full length protein (although ligation can also be used to make libraries of partial nucleic acid sequences which can then be recombined, e.g., to produce a partial or full-length recombinant nucleic acid). The oligonucleotide set typically includes at least a first oligonucleotide which is complementary to at least a first of the parental nucleic acids at a first region of sequence diversity and at least a second oligonucleotide which is complementary to at least a second of

the parent nucleic acids at a second region of diversity. The parent nucleic acids can be homologous or non-homologous. Often, nucleic acids such as oligos are ligated with a ligase. In one typical format, oligonucleotides are hybridized to a first parent nucleic acid which acts as a template, and ligated with a ligase. The oligos can also be extended with a polymerase and ligated. The polymerase can be, e.g., an ordinary DNA polymerase or a thermostable DNA polymerase. The ligase can also be an ordinary DNA ligase, or a thermostable DNA ligase. In one set of approaches, a common element for non-PCR based recombination methods is preparation of a single-stranded template to which primers are annealed and then elongated by a DNA polymerase in the presence of dNTP's and R408 to package ss plasmids into filamentous phage such as M13KO7 (Pharmacia Biotech) or region into a plasmid and use of a helper phage such as M13KO7 (Pharmacia Biotech) or Dalle et al. (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" Methods Mol Biol. 57:369-74; Smith (1985) "In vitro mutagenesis" Annu Rev Genet. 19:423-462; Bostein & Shorlie (1985) "Strategies and applications of *in vitro* mutagenesis" Science 229:1193-1201; and Carter (1986) "Site-directed mutagenesis" Biocchem J. 237:1-7; Kunikei (1987) "The efficiency of oligonucleotide directed mutagenesis" Nucleic Acids & Molecular Biology (1987); Eckstein, F. and Lillie, D.M.J. eds Springer Verlag, Berlin.

30 For example, in one aspect, a "Kunkel style" method uses uracil containing templates. Similarly, the "Eckstein" method uses phosphorothioate-modified DNA (Taylor et al. (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA." Nucleic Acids Res. 13:8749-8764; Taylor et al. (1985) "The rapid preparation of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA." Nucleic Acids & Molecular Biology (1987); Eckstein, F. and Lillie, D.M.J. eds Springer Verlag, Berlin.

25 For example, the "Kunkel style" method uses uracil containing templates can be performed using available methods. See e.g., Ling et al. (1997) "Approaches to DNA mutagenesis: an overview." Annu Biocchem. Dec 15;254(2):157-78; Dalle et al. (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" Methods Mol Biol. 57:369-74; Smith (1985) "In vitro mutagenesis" Annu Rev Genet. 19:423-462; Bostein & Shorlie (1985) "Strategies and applications of *in vitro* mutagenesis" Science 229:1193-1201; and Carter (1986) "Site-directed mutagenesis" Biocchem J. 237:1-7; Kunikei (1987) "The efficiency of oligonucleotide directed mutagenesis" Nucleic Acids & Molecular Biology (1987); Eckstein, F. and Lillie, D.M.J. eds Springer Verlag, Berlin.

20 For example, the ss template can be prepared by incorporation of the phage IgG region into a plasmid and use of a helper phage such as M13KO7 (Pharmacia Biotech) or R408 to package ss plasmids into filamentous phage such as M13KO7 (Pharmacia Biotech) or region into a plasmid and use of a helper phage such as M13KO7 (Pharmacia Biotech) or Dalle et al. (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" Methods Mol Biol. 57:369-74; Smith (1985) "In vitro mutagenesis" Annu Rev Genet. 19:423-462; Bostein & Shorlie (1985) "Strategies and applications of *in vitro* mutagenesis" Science 229:1193-1201; and Carter (1986) "Site-directed mutagenesis" Biocchem J. 237:1-7; Kunikei (1987) "The efficiency of oligonucleotide directed mutagenesis" Nucleic Acids & Molecular Biology (1987); Eckstein, F. and Lillie, D.M.J. eds Springer Verlag, Berlin.

15 For example, the ss template can be prepared by incorporation of the phage IgG region into a plasmid and use of a helper phage such as M13KO7 (Pharmacia Biotech) or R408 to package ss plasmids into filamentous phage such as M13KO7 (Pharmacia Biotech) or region into a plasmid and use of a helper phage such as M13KO7 (Pharmacia Biotech) or Dalle et al. (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" Methods Mol Biol. 57:369-74; Smith (1985) "In vitro mutagenesis" Annu Rev Genet. 19:423-462; Bostein & Shorlie (1985) "Strategies and applications of *in vitro* mutagenesis" Science 229:1193-1201; and Carter (1986) "Site-directed mutagenesis" Biocchem J. 237:1-7; Kunikei (1987) "The efficiency of oligonucleotide directed mutagenesis" Nucleic Acids & Molecular Biology (1987); Eckstein, F. and Lillie, D.M.J. eds Springer Verlag, Berlin.

10 In one set of approaches, a common element for non-PCR based recombination methods is preparation of a single-stranded template to which primers are annealed and then elongated by a DNA polymerase in the presence of dNTP's and R408 to package ss plasmids into filamentous phage such as M13KO7 (Pharmacia Biotech) or region into a plasmid and use of a helper phage such as M13KO7 (Pharmacia Biotech) or Dalle et al. (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" Methods Mol Biol. 57:369-74; Smith (1985) "In vitro mutagenesis" Annu Rev Genet. 19:423-462; Bostein & Shorlie (1985) "Strategies and applications of *in vitro* mutagenesis" Science 229:1193-1201; and Carter (1986) "Site-directed mutagenesis" Biocchem J. 237:1-7; Kunikei (1987) "The efficiency of oligonucleotide directed mutagenesis" Nucleic Acids & Molecular Biology (1987); Eckstein, F. and Lillie, D.M.J. eds Springer Verlag, Berlin.

5 Often, nucleic acids such as oligos are ligated with a ligase. In one typical format, oligonucleotides are hybridized to a first parent nucleic acid which acts as a template, and ligated with a ligase. The oligos can also be extended with a polymerase and ligated. The polymerase can be, e.g., an ordinary DNA polymerase or a thermostable DNA polymerase. The ligase can also be an ordinary DNA ligase, or a thermostable DNA ligase. In one set of approaches, a common element for non-PCR based recombination methods is preparation of a single-stranded template to which primers are annealed and then elongated by a DNA polymerase in the presence of dNTP's and R408 to package ss plasmids into filamentous phage such as M13KO7 (Pharmacia Biotech) or region into a plasmid and use of a helper phage such as M13KO7 (Pharmacia Biotech) or Dalle et al. (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" Methods Mol Biol. 57:369-74; Smith (1985) "In vitro mutagenesis" Annu Rev Genet. 19:423-462; Bostein & Shorlie (1985) "Strategies and applications of *in vitro* mutagenesis" Science 229:1193-1201; and Carter (1986) "Site-directed mutagenesis" Biocchem J. 237:1-7; Kunikei (1987) "The efficiency of oligonucleotide directed mutagenesis" Nucleic Acids & Molecular Biology (1987); Eckstein, F. and Lillie, D.M.J. eds Springer Verlag, Berlin.

generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" Nucleic Acids Res. 13:8765-8787; Nakamaye & Eckstein (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis." Nucleic Acids Res. 14: 9679-9698; 5 Sayers et al. (1988). "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis." Nucleic Acids Res. 16:791-802; Sayers et al. (1988) "5'-3' Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" Nucleic Acids Res. 16:803-814). The use of restriction selection, or e.g., purification can be used in conjunction with mismatch repair deficient 10 strains (see, e.g., Carter et al. (1985) "Improved oligonucleotide site directed mutagenesis using M13 vectors" Nucleic Acids Res. 13, 4431-4443 Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors." Methods in Enzymol. 154:382-403; Wells (1986) "Importance of hydrogen bond formation in stabilizing the transition state of subtilisin." Trans. R. Soc. Lond. A317, 415-423).

15 The "mutagenic" primer used in these methods can be a synthetic oligonucleotide encoding any type of randomization, insertion, deletion, family gene shuffling oligonucleotide based on sequence diversity of homologous genes, etc. The primer(s) could also be fragments of homologous genes that are annealed to the ss parent template. In this way chimeras between 2 or more parental genes can be generated.

20 Multiple primers can anneal to a given template and be extended to create multiply chimeric genes. The use of a DNA polymerase such as those from phages T4 or T7 are suitable for this purpose as they do not degrade or displace a downstream primer from the template.

25 For example, in one aspect, DNA shuffling is performed using uracil containing templates. In this embodiment, the gene of interest is cloned into an *E. coli* plasmid containing the filamentous phage intergenic (IG, ori) region. Single stranded (ss) plasmid DNA is packaged into phage particles upon infection with a helper phage such as M13KO7 (Pharmacia) or R408 and can be easily purified by methods such as phenol/chloroform extraction and ethanol precipitation. If this DNA is prepared in a dut-
30 ung- strain of *E. coli*, a small number of uracil residues are incorporated into it in place of the normal thymine residues. One or more primers or other oligos as described above are annealed to the ss uracil-containing template by heating to 90°C and slowly cooling to room temperature. An appropriate buffer containing all 4 deoxyribonucleotides, T7 DNA

are found in Paten and Stemmer, USSN 60/102,362 "SHUFFLING OF CODON ALTERED ADDITIONAL DETAILS ON CODON MODIFICATION PROCEDURES AND THEIR APPLICATION TO DNA SHUFFLING". Additional details on codon modification procedures and their application to DNA shuffling which, in turn, increases the frequency of recombination between the nucleic acids, used in particular positions, it is possible to increase the similarity between the nucleic acids, 30 distantly related nucleic acids which are to be recombined. By selecting codons are expressed, codon preference can also be used to increase sequence similarity between expression to the selection of oligonucleotide sequences to optimize described *supra*, are optionally used to modify codons in synthetic protocols.

simply by reference to well-known codon-bias tables. Codon-based synthetic methods, as 25 strategies herein typically are made synthetically, selecting optimal codon preference is done expression of the recombinant nucleic acid. Because the oligonucleotides used in the various particular cell in which selection is to take place typically results in maximization of assessed or otherwise selected. Conforming a recombinant nucleic acid to the codon bias of a 20 cell in which a recombinant product of an oligonucleotide shuffling procedure is to be derived. In particular, it is useful, e.g., to modify codon preference to optimize expression in codon use as compared to the parental sequences from which the oligonucleotides are 15 used.

In one aspect, the oligonucleotides utilized in the methods herein have altered chimeric genes can be generated.

CODON MODIFICATION

primers are derived from or correspond to fragments of homologous genes, then multiply then it is possible to simultaneously introduce numerous changes in a single reaction. If the predominantly contain the desired change in the gene of interest. If multiple primers are used will correct it by using the newly synthesized strand as a template. The resulting plasmids removes it, creating asymmetric sites that are either not replicated or the host repair systems (ungene product) enzyme in the host cell recognizes the uracil in the template strand and 10 strain of E. coli and antibiotic selection for the plasmid is applied. The uracil-N-glycosylase primer, stop and ligase will seal the gap. This reaction is then transformed into an ung+ If multiple primers are used, then the polymerase will extend to the next synthesized strand and the 5' end of the primer.

incorporating the primer. DNA ligase seals the gap between the 3' end of the newly extended from the 3' end of the primer and synthesizes a complementary strand to the template between room temperature and e.g., about 37°C for ≥1 hour. The T7 DNA polymerase polymerase and T4 DNA ligase is added to the annealed template/primer mix and incubated

NUCLEIC ACIDS," filed September 29, 1998 and related application of Paten and Stemmer, Attorney docket number 018097-028510, entitled "SHUFFLING OF CODON ALTERED NUCLEIC ACIDS," filed January 29, 1999.

LENGTH VARIATION BY MODULAR SHUFFLING

5 Many functional sequence domains for genes and gene elements are composed of functional subsequence domains. For example, promoter sequences are made up of a number of functional sequence elements which bind transcription factors, which, in turn, regulate gene expression. Enhancer elements can be combined with promoter elements to enhance expression of a given gene. Similarly, at least some exons represent modular
10 domains of an encoded protein, and exons can be multimerized or deleted relative to a wild-type gene and the resulting nucleic acids recombined to provide libraries of altered gene (or encoded protein) modules (i.e., libraries of module inserted or deleted nucleic acids). The number and arrangement of modular sequences, as well as their sequence composition, can affect the overall activity of the promoter, exon, or other genetic module.

15 The concept of exons as modules of genes and encoded proteins is established, particularly for proteins which have developed in eukaryotes. *See, e.g.,* Gilbert and Glynias (1993) *Gene* 137-144; Dolittle and Bork (October 1993) *Scientific American* 50-56; and Patthy (1991) *Current Opinions in Structural Biology* 1:351-361. Shuffling of exon modules is optimized by an understanding of exon shuffling rules. Introns (and consequently exons)
20 occur in three different phases, depending on the splice junction of a codon at the exon-intron boundary. *See, Stemmer (1995) Biotechnology* 13:549-553; Patthy (1994) *Current Opinions in Structural Biology* 4:383-392 and Patthy (1991) *Current Opinions in Structural Biology* 1:351-361.

25 In nature, splice junctions of shuffled exons have to be "phase compatible" with those of neighboring exons—if not, then a shift in reading frame occurs, eliminating the information of the exon module. The three possible phases of an intron are phases 1, 2, or 0, for the base position within the codon at the intron-exon boundary in which the intron occurs. Classification of introns according to their location relative to the reading frame is as follows:
30 a phase 0 intron is located between two codons of flanking exons; a phase 1 intron is located between the first and second nucleotide of a codon and a phase 2 intron is located between the second and third nucleotide of a codon. Phase 1 introns are the most common in nature.

One aspect of the present invention is the shuffling of modular sequences (including, e.g., promoter elements and exons) to vary the sequence of such modules, the

phase (i.e., having the same reading frame) to increase the likelihood that any given library the recombination methods are typically selected to result in exons being joined in the same homologous parental nucleic acid. When the modules are exons, the oligonucleotides used in in the number of copies and arrangement of a given module relative to a corresponding or 30 libraries resulting from module deletion/insertion strategies noted above vary

elongation of a mixture of oligonucleotides, and the like.

oligonucleotides (and, optionally, of associated sequences) during hybridization and with ends that have regions of identity permitting multimerization of the one or more sequences and/or insertion of a module in a selected position), one or more oligonucleotides 25 corresponding to more than one sequence module (providing for deletion of intervening sequences. Examples of such oligonucleotides include those having subsequences modules. Oligonucleotides which provide for insertion, deletion or multimerization of sequence activity of a selected nucleic acid, while module shuffling oligonucleotides are

subsequences of a given nucleic acid which provide an activity or distinct component of an herein to recombine homologous nucleic acids. Thus, sequence module elements are those "shuffling" oligonucleotides can be used in conjunction with any of the other approaches which provide for multimerization of sequence module elements. These "module 20

of overlapping family gene shuffling oligonucleotides is constructed to comprise oligos multimers or nucleic acids comprising module deletions separately. Thus, in one aspect, a set oligonucleotide recombination and gene assembly, thereby avoiding the need to make 15

selected modules such as exons, promoter elements, enhancers, or the like during In a preferred aspect, oligos are selected to generate multimers and/or to delete

gene reassembly/ recombination reaction as discussed *supra*.
the modules to be multimerized. The multimers can be made separately, or can be oligos in multimers, with the overall lengths of such multimers being dependent on the molar ratios of incorporated into final shuffled products by homologous recombination at the ends of the creation of multimers in a subsequent recombination reaction. These multimers can be is spiked into a recombination reaction, with ends designed to be complementary, causing the For example, a PCR-generated fragment containing the element to be repeated 10

oligonucleotides into a recombination mixture.

length variation approaches simply by spiking appropriately designed fragments or oligonucleotide-mediated methods herein, can be combined with element duplication and copies) and the length of the modules. In particular, standard shuffling methods, and/or the number of repeats of modules (from 0 (i.e., a deletion of the element) to a desired number of 5

member will be functionally active. This is illustrated schematically in Figure 3. The differently shaded modules represent separate exons, with the phase of the exon being indicated as 1, 2, or 0.

SHUFFLING OF CLADISTIC INTERMEDIATES

5 The present invention provides for the shuffling of "evolutionary intermediates." In the context of the present invention, evolutionary intermediates are artificial constructs which are intermediate in character between two or more homologous sequences, e.g., when the sequences are grouped in an evolutionary dendogram.

10 Nucleic acids are often classified into evolutionary dendograms (or "trees") showing evolutionary branch points and, optionally, relatedness. For example, cladistic analysis is a classification method in which organisms or traits (including nucleic acid or polypeptide sequences) are ordered and ranked on a basis that reflects origin from a postulated common ancestor (an intermediate form of the divergent traits or organisms). Cladistic analysis is primarily concerned with the branching of relatedness trees (or 15 "dendograms") which shows relatedness, although the degree of difference can also be assessed (a distinction is sometimes made between evolutionary taxonomists who consider degrees of difference and those who simply determine branch points in an evolutionary dendogram (classical cladistic analysis); for purposes of the present invention, however, relatedness trees produced by either method can produce evolutionary intermediates).

20 Cladistic or other evolutionary intermediates can be determined by selecting nucleic acids which are intermediate in sequence between two or more extant nucleic acids. Although the sequence may not exist in nature, it still represents a sequence which is similar to a sequence in nature which had been selected for, i.e., an intermediate of two or more sequences represents a sequence similar to the common ancestor of the two or more extant 25 nucleic acids. Thus, evolutionary intermediates are one preferred shuffling substrate, as they represent "pseudo selected" sequences, which are more likely than randomly selected sequences to have activity.

30 One benefit of using evolutionary intermediates as substrates for shuffling (or of using oligonucleotides which correspond to such sequences) is that considerable sequence diversity can be represented in fewer starting substrates (i.e., if starting with parents A and B, a single intermediate "C" has at least a partial representation of both A and B). This simplifies the oligonucleotide synthesis for gene reconstruction/ recombination methods, improving the efficiency of the procedure. Further, searching sequence databases with

Family shuffling of genes is a good way to access functional diversity of encoded proteins. It can be advantageous, however, to shuffle only a portion of an encoded protein which provides an activity of interest, particularly where the protein is multifunctional and one or more activity can be mapped to a subsequence (a domain) of the overall protein.

For example, enzymes such as glycosyl transferases have two substrates: the acceptor and the activated sugar donor. To change the sugar to be transferred without altering the acceptor, it can be preferable to family shuffle only the sugar binding domain, since family shuffling the sugar acceptor domain can result in lowered numbers of the desired

PROTEIN DOMAIN SHUFFLING

Intermediate sequences can also be selected between two or more synthetic sequences which are not represented in nature, simply by starting from two synthetic sequences. Such synthetic sequences can include evolutionary intermediates, proposed gene sequences, or other sequences of interest that are related by sequence. These "artificial intermediates" are also useful in reducing the complexity of gene reconstruction methods and for improving the ability to search evolutionary databases.

evolutionarily intermediate microRNAs increases the chances of identifying related nucleic acids using standard search programs such as BLAST.

In one example, there are 5 enzymes, eA-eE (each of 500 amino acids) that transfers sugars a-e to acceptors A-E. To generate a library of enzymes that transfer sugars a-e to acceptor A it can be preferable to shuffling the sugar binding domains of eA-eE, combining them with acceptor binding domains of eA.

5 One technical challenge in practicing this strategy is that there can be insufficient data to identify such functional domains in a protein of interest. When this is the case, a set of libraries can be generated by family shuffling random portions of the enzyme. For example, as applied to the family shuffling of enzymes eA-eE, *above*, a first library can be made encoding the first 100 amino acids of eA-eE, in combination with the last 400 amino 10 acids of any one of eA-eE by appropriately selecting oligonucleotide sets for recombination and elongation. A second library can be made which family shuffles the second 100 amino acids of eA-eE, in combination with encoding the first 100 amino acids of any one of eA-Ae and the last 300 amino acids of any one of eA-Ae, and so on. Small subsets of these libraries are screened for a first desired function. Libraries that have retained the first desired function 15 (e.g., acceptor activity in the example above) have a relatively higher proportion of variants in additional selectable functions (e.g., sugar transfer in the example above).

This approach can be used for diversification of any multi-functional protein in which one property is desirably conserved. This strategy is particularly advantageous when the property to be conserved is complex (e.g., substrate specificity for, e.g., 20 polyketides, non-ribosomal peptides or other natural products).

In general, selection of oligonucleotides to provide shuffling of individual domains (whether corresponding to known functional subsequences or to subsequences of unknown function as noted above) is performed by providing two general types of sequence-related oligonucleotides. The first type is provided by selecting sequence-related overlapping 25 oligonucleotide sets corresponding to regions where recombination is desired (i.e., according to the strategies noted herein), while the second type provides recombination junctions between the domains to be shuffled and non-shuffled domains, i.e., similar to a crossover oligonucleotide as described herein. The non-shuffled domains can be produced by simple oligonucleotide gene reconstruction methods (e.g., using ligation or polymerase-mediated 30 extension reactions to concatenate oligonucleotides), or the non-shuffled domains can be produced by enzymatic cleavage of larger nucleic acids.

capture these non-natural (but still predicted to be active) homologues. A variety of protein with little or no change in function. Sets of family shuffling oligos are then made which the protein interacts with a ligand), regions can be defined which are predicted to be mutable Second, where structural information is available for a protein (and, e.g., how

factor (G-CSF).

Protein Science 6:1228-1236 describe alanine scanning of granulocyte colony stimulating proteins, alanine scanning information is already available. For example, Young et al. (1997)

capture these non-natural (but still active) homologues. For many commercially important

essentially neutral with respect to function. Sets of shuffling oligos are then made which acids could be tested at each position to define a large spectrum of point mutations that are

positions that can be mutated without a significant loss of function. In principle, all 20 amino

First, point mutagenesis (e.g., alanine scanning) can be performed to define

(protein, RNA, etc.). At least two basic approaches to can be used.

gene without significantly degrading the desired functional property of an encoded molecule undesirable sequence space is to define those positions which can be modified in a given

One way of capturing non-natural diversity without significantly increasing

overall potential sequence space in which recombination occurs.

natural diversity represents evolutionarily tested molecules, representing a more targeted space is undesirable "noise" which has to be selected out of the population. In contrast,

characteristic. For these homologues, much of the resulting increase in selectable sequence that many of the resulting homologues are not functional with respect to a relevant

natural diversity.

functional properties that are not provided by recombination of nucleic acids representing nucleic acids. This additional diversity allows for the development or acquisition of

other than naturally occurring sequence space is accessed by the resulting recombination.

One advantage of recombining non-natural homologues is that sequence space nucleic acids can be natural or non-natural (i.e., artificial) homologues.

gene synthesis and recombination protocols as discussed *supra*. As noted, the homologous nucleic acids by making sets of family shuffling oligonucleotides which are recombined in

Familly based oligo shuffling involves the recombination of homologous

AND ALANINE SCANNING

EXPANDED FAMILY SHUFFLING INCORPORATING MOLECULAR MODELING

crystal structures are available (including, e.g., the crystal structure of G-CSF: Hill et al. (1993) *PNAS* 90:5167).

Similarly, even where structural information is not available, molecular modeling can be performed to provide a predicted structure, which can also be used to predict which residues can be changed without altering function. A variety of protein structure modeling programs are commercially available for predicting protein structure. Further, the relative tendencies of amino acids to form regions of superstructure (helices, β -sheets, etc.) are well established. For example, O'Neil and DeGrado *Science* v.250 provide a discussion of the helix forming tendencies of the commonly occurring amino acids. Tables of relative structure forming activity for amino acids can be used as substitution tables to predict which residues can be functionally substituted in a given portion. Sets of family shuffling oligos are then made which capture these non-natural (but still predicted to be active) homologues.

For example, Protein Design Automation (PDA) is one computationally driven system for the design and optimization of proteins and peptides, as well as for the design of proteins and peptides. Typically, PDA starts with a protein backbone structure and designs the amino acid sequence to modify the protein's properties, while maintaining its three dimensional folding properties. Large numbers of sequences can be manipulated using PDA, allowing for the design of protein structures (sequences, subsequences, etc.). PDA is described in a number of publications, including, e.g., Malakauskas and Mayo (1998) "Design, Structure and Stability of a Hyperthermophilic Protein Variant" *Nature Struc. Biol.* 5:470; Dahiyat and Mayo (1997) "De Novo Protein Design: Fully Automated Sequence Selection" *Science*, 278, 82-87. DeGrado, (1997) "Proteins from Scratch" *Science*, 278:80-81; Dahiyat, Sarisky and Mayo (1997) "De Novo Protein Design: Towards Fully Automated Sequence Selection" *J. Mol. Biol.* 273:789-796; Dahiyat and Mayo (1997) "Probing the Role of Packing Specificity in Protein Design" *Proc. Natl. Acad. Sci. USA*, 94:10172-10177; Hellinga (1997) "Rational Protein Design – Combining Theory and Experiment" *Proc. Natl. Acad. Sci. USA*, 94:10015-10017; Su and Mayo (1997) "Coupling Backbone Flexibility and Amino Acid Sequence Selection in Protein Design" *Prot. Sci.* 6:1701-1707; Dahiyat, Gordon and Mayo (1997) "Automated Design of the Surface Positions of Protein Helices" *Prot. Sci.*, 6:1333-1337; Dahiyat and Mayo (1996) "Protein Design Automation" *Prot. Sci.*, 5:895-903. Additional details regarding PDA are available, e.g., at <http://www.xencor.com/>. PDA can be used to identify variants of a sequence that are likely to retain activity, providing a set of

FOSI-RECOMBINANT SCREENING TECHNOLOGIES

homologous nucleic acids that can be used as a basis for oligonucleotide mediated recombination.

WO 00/422581 PCT/US00/01263

Depending on the particular screening protocol used for a desired property, initial round(s) of screening can sometimes be performed using bacterial cells due to high transfection efficiencies and ease of culture. However, bacterial expression is often not practical or desired, and yeast, fungal or other eukaryotic systems are also used for library

The nature of screening or selection depends on what property or characteristic is to be acquired or the property or characteristic for which improvement is sought, and many examples are discussed below. It is not usually necessary to understand the molecular basis by which particular products of recombination (recombinant segments) have acquired new or improved properties or characteristics relative to the starting substrates. For example, a gene can have many component sequences, each having a different intended role (e.g., coding sequence, regulatory sequences, targeting sequences, stability-conferring sequences, subunit sequences and sequences affecting integration). Each of these component sequences can be varied and recombined simultaneously. Screening/selection can then be performed, for example, for recombinant segments that have increased ability to confer activity upon a cell without the need to attribute such improvement to any of the individual sequences.

The precise screening method that is used in the various shuffling procedures herein is not a critical aspect of the invention. In general, one of skill can practice appropriate screening (i.e., selection) methods, by reference to the activity to be selected for. In any case, one or more recombination cycle(s) is/are usually followed by one or more cycle of screening or selection for molecules or transformed cells or organisms having a desired property, trait or characteristic. If a recombination cycle is performed *in vitro*, the products of recombination, i.e., recombinant segments, are sometimes introduced *in vitro* cells before the screening step. Recombinant segments can also be linked to an appropriate vector or other regulatory sequences before screening. Alternatively, products of recombination generated *in vitro* are sometimes packaged in viruses (e.g., bacteriophage) before screening. If recombination is performed *in vivo*, recombination products can sometimes be screened in the cells in which recombination occurred. In other applications, sometimes segmental segments are extracted from the cells, and optionally packaged as viruses, before screening.

POST-RECOMBINATION SCREENING TECHNIQUES

expression and screening. Similarly, other types of screening which are not amenable to screening in bacterial or simple eukaryotic library cells, are performed in cells selected for use in an environment close to that of their intended use. Final rounds of screening can be performed in the precise cell type of intended use.

5 One approach to screening diverse libraries is to use a massively parallel solid-phase procedure to screen shuffled nucleic acid products, e.g., encoded enzymes, for enhanced activity. Massively parallel solid-phase screening apparatus using absorption, fluorescence, or FRET are available. *See, e.g.,* United States Patent 5,914,245 to Bylina, et al. (1999); *see also,* <http://www.kairos-scientific.com/>; Youvan et al. (1999) "Fluorescence 10 Imaging Micro-Spectrophotometer (FIMS)" Biotechnology et alia<www.et-al.com> 1:1-16; Yang et al. (1998) "High Resolution Imaging Microscope (HIRIM)" Biotechnology et alia, <www.et-al.com> 4:1-20; and Youvan et al. (1999) "Calibration of Fluorescence Resonance 15 Energy Transfer in Microscopy Using Genetically Engineered GFP Derivatives on Nickel Chelating Beads" posted at www.kairos-scientific.com. Following screening by these techniques, sequences of interest are typically isolated, optionally sequenced and the sequences used as set forth herein to design new oligonucleotide shuffling methods.

20 If further improvement in a property is desired, at least one and usually a collection of recombinant segments surviving a first round of screening/selection are subject to a further round of recombination. These recombinant segments can be recombined with each other or with exogenous segments representing the original substrates or further variants thereof. Again, recombination can proceed *in vitro* or *in vivo*. If the previous screening step identifies desired recombinant segments as components of cells, the components can be subjected to further recombination *in vivo*, or can be subjected to further recombination *in vitro*, or can be isolated before performing a round of *in vitro* recombination. Conversely, if 25 the previous screening step identifies desired recombinant segments in naked form or as components of viruses, these segments can be introduced into cells to perform a round of *in vivo* recombination. The second round of recombination, irrespective how performed, generates further recombinant segments which encompass additional diversity than is present in recombinant segments resulting from previous rounds.

30 The second round of recombination can be followed by a further round of screening/selection according to the principles discussed above for the first round. The stringency of screening/selection can be increased between rounds. Also, the nature of the screen and the property being screened for can vary between rounds if improvement in more than one property is desired or if acquiring more than one new property is desired.

Additional rounds of recombination and screening can then be performed until the recombinant segments have sufficiently evolved to acquire the desired new or improved property or function.

The nucleic acids produced by the methods of the invention are optionally cloned into cells for activity screening (or used in *in vitro* transcription reactions to make products which are screened). Furthermore, the nucleic acids can be sequenced, expressed, amplified in *in vitro* or treated in any other common recombinant method.

General texts which describe molecular biological techniques useful herein, including cloning, mutagenesis, library construction, screening assays, cell culture and the like include Berger and Kimmel, *Guide to Molecular Cloning Techniques*, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a Laboratory Manual (Berger and Kimmel), *Guide to Molecular Cloning Techniques*, Methods in Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., supplemented throughout with nucleic acids. In addition to Berger, Ausubel and Sambrook, useful general references for nucleic acids are generally available, as are methods of expressing proteins encoded by such nucleic acids. In addition to Berger, Ausubel and Sambrook, useful general references for culture of animal cells include Freshney (Culture of Animal Cells, A Manual of Basic Cell Cloning, culture and regeneration include Payne et al. (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. New York, NY (Payne); and Gamborg and Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture: Fundamental Methods* Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) (Gamborg). A variety of Cell culture media are described in Atlas and Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, FL (Atlas). Additional information for plant cell culture is found in available commercial literature such as the *Life Science Research Cell Culture Catalogue* (1998) from Sigma-Aldrich, Inc (St Louis, MO) (Sigma-LSRCC) and, e.g., the *Plant Culture Catalogue and Supplement* (1997) also from Sigma-Aldrich, Inc (St Louis, MO) (Sigma).

POST-SHUFFLING PROCEDURES

Additional rounds of recombination and screening can then be performed until the recombinant segments have sufficiently evolved to acquire the desired new or improved property or function.

Additional rounds of recombination and screening can then be performed until the recombinant segments have sufficiently evolved to acquire the desired new or improved property or function.

Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods, useful e.g., for amplifying oligonucleotide shuffled nucleic acids including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Q β -replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA). These 5 techniques are found in Berger, Sambrook, and Ausubel, *id.*, as well as in Mullis *et al.*, (1987) U.S. Patent No. 4,683,202; PCR Protocols A Guide to Methods and Applications (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) C&EN 36-47; The Journal Of NIH Research (1991) 3, 81-94; Kwok *et al.* 10 (1989) Proc. Natl. Acad. Sci. USA 86, 1173; Guatelli *et al.* (1990) Proc. Natl. Acad. Sci. USA 87, 1874; Lomell *et al.* (1989) J. Clin. Chem 35, 1826; Landegren *et al.*, (1988) Science 241, 1077-1080; Van Brunt (1990) Biotechnology 8, 291-294; Wu and Wallace, (1989) Gene 4, 560; Barringer *et al.* (1990) Gene 89, 117, and Sooknanan and Malek (1995) 15 Biotechnology 13: 563-564. Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No. 5,426,039. Improved methods of amplifying large nucleic acids by PCR are summarized in Cheng *et al.* (1994) Nature 369: 684-685 and 20 the references therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. See, Ausubel, Sambrook and Berger, *all supra*. In one preferred method, reassembled sequences are checked for incorporation of family gene shuffling oligonucleotides. This can be done by cloning and sequencing the nucleic acids, and/or by 25 restriction digestion, e.g., as essentially taught in Sambrook, Berger and Ausubel, *above*. In addition, sequences can be PCR amplified and sequenced directly. Thus, in addition to, e.g., Sambrook, Berger, Ausubel and Innis (*id.* and *above*), additional PCR sequencing PCR 30 methodologies are also particularly useful. For example, direct sequencing of PCR generated amplicons by selectively incorporating boronated nuclease resistant nucleotides into the amplicons during PCR and digestion of the amplicons with a nuclease to produce sized template fragments has been performed (Porter *et al.* (1997) Nucleic Acids Research 25(8):1611-1617). In the methods, 4 PCR reactions on a template are performed, in each of which one of the nucleotide triphosphates in the PCR reaction mixture is partially substituted with a 2' deoxynucleoside 5'-[P-borano]-triphosphate. The boronated nucleotide is stochastically incorporated into PCR products at varying positions along the PCR amplicon in a nested set of PCR fragments of the template. An exonuclease which is blocked by incorporated boronated nucleotides is used to cleave the PCR amplicons. The cleaved

effects are to be observed, every 5-10 wells can test a single variant. Thus, a single standard microtiter plate can be used to run a separate assay, or, if concentration or incubation time several thousand different shuffled variants in a single day. In particular, each well of a 30 assay to screen up to several components, cell or organismal viability, production of reporter compositions, and the like. proteins, receptors, ligands and the like are known. Formats include binding to immobilized Of course, the relevant assay will depend on the application. Many assays for such assays to oligonucleotide selection, synthesis and recombination.

25 components can provide for high-throughput screening assays, as well as the coupling of physical oligonucleotide synthesis. In addition, other important integrated system computers having appropriate software can be used to perform "in silico" shuffling prior to alignment of nucleic acids using a computer and sequence alignment software. Similarly, As noted throughout, one preferred aspect of the present invention is the 20 INTEGRATED ASSAYS AND INTEGRATED SYSTEM ELEMENTS

produce corresponding molecules, e.g., by oligonucleotide synthesis and reassembly PCR. predicting recombinational outcomes. The predicted recombinational outcomes are used to manual inspection and alignment) such as those representing homologous nucleic acids and acid, e.g., by aligning nucleic acid sequence strings (using standard alignment software, or by used to model recombinational or mutational events which can occur in one or more nucleic such as point mutations, recombination of two strands of homologous nucleic acids, etc.) are 15 In brief, genetic operators (algorithms which represent given genetic events herewith.

CHARACTERISTICS" by Seifonov et al., attorney docket number 02-289-3US, filed POLYNUCLEOTIDES AND POLYPEPTIDES HAVING DESIRED

10 described in detail in "METHODS FOR MAKING CHARACTER STRINGS, by reassembly PCR of synthetic oligonucleotides as described herein. In silico shuffling is sequence strings are combined in a computer system and desirable products are made, e.g., sequence strings are shuffled in a computer system and desirable products are made, e.g., shuffling using genetic operators in a computer. As applied to the present invention, gene 5 "in silico" shuffling utilizes computer algorithms to perform "virtual"

IN SILICO SHUFFLING

manipulations than performing standard Sanger-style sequencing of PCR amplicons. sequence of the amplicon. An advantage of this method is that it uses fewer biochemical manipulations are then separated by size using polyacrylamide gel electrophoresis, providing the

microtiter plate can assay about 100 (e.g., 96) reactions. If 1536 well plates are used, then a single plate can easily assay from about 100- about 1500 different reactions. It is possible to assay several different plates per day; assay screens for up to about 6,000-20,000 different assays (i.e., involving different nucleic acids, encoded proteins, concentrations, etc.) is 5 possible using the integrated systems of the invention. More recently, microfluidic approaches to reagent manipulation have been developed, e.g., by Caliper Technologies (Mountain View, CA).

In one aspect, library members, e.g., cells, viral plaques, spores or the like, are separated on solid media to produce individual colonies (or plaques). Using an automated 10 colony picker (e.g., the Q-bot, Genetix, U.K.), colonies or plaques are identified, picked, and up to 10,000 different mutants inoculated into 96 well microtiter dishes containing two 3 mm glass balls/well. The Q-bot does not pick an entire colony but rather inserts a pin through the center of the colony and exits with a small sampling of cells, (or mycelia) and spores (or viruses in plaque applications). The time the pin is in the colony, the number of dips to 15 inoculate the culture medium, and the time the pin is in that medium each effect inoculum size, and each can be controlled and optimized. The uniform process of the Q-bot decreases human handling error and increases the rate of establishing cultures (roughly 10,000/4 hours). These cultures are then shaken in a temperature and humidity controlled incubator. The glass 20 balls in the microtiter plates act to promote uniform aeration of cells and the dispersal of mycelial fragments similar to the blades of a fermenter. Clones from cultures of interest can be cloned by limiting dilution. As also described supra, plaques or cells constituting libraries can also be screened directly for production of proteins, either by detecting hybridization, protein activity, protein binding to antibodies, or the like.

A number of well known robotic systems have also been developed for 25 solution phase chemistries useful in assay systems. These systems include automated workstations like the automated synthesis apparatus developed by Takeda Chemical Industries, LTD. (Osaka, Japan) and many robotic systems utilizing robotic arms (Zymate II, Zymark Corporation, Hopkinton, Mass.; Orca, Beckman Coulter, Inc. (Fullerton, CA)) which mimic the manual synthetic operations performed by a scientist. Any of the above devices 30 are suitable for use with the present invention, e.g., for high-throughput screening of molecules assembled from the various oligonucleotide sets described herein. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein with reference to the integrated system will be apparent to persons skilled in the relevant art.

High throughput screening systems are commercially available (see, e.g., Zymark Corp., Hopkinton, MA; Air Technical Industries, Mentor, OH; Beckman Instruments, Inc. Fullerton, CA; Precision Systems, Inc., Natick, MA, etc.). These systems dispense liquid reagent(s) appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. The manufacturers of such systems provide detailed protocols for the various high throughput. Thus, for example, Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like.

assay component. The image scanner interfaces with the image analysis software to provide a measurement of probe label intensity.

Of course, these assay systems can also include integrated systems incorporating oligonucleotide selection elements, such as a computer, database with nucleic acid sequences of interest, sequence alignment software, and oligonucleotide selection software. In addition, this software can include components for ordering the selected oligonucleotides, and/or directing synthesis of oligonucleotides by an operably linked oligonucleotide synthesis machine. Thus, the integrated system elements of the invention 5 optionally include any of the above components to facilitate high throughput recombination and selection. It will be appreciated that these high-throughput recombination elements can be in systems separate from those for performing selection assays, or the two can be 10 integrated.

In one aspect, the present invention comprises a computer or computer readable medium with an instruction set for selecting an oligonucleotide set such as a set of 15 family shuffling oligonucleotides using the methods described herein. The instruction set aligns homologous nucleic acids to identify regions of similarity and regions of diversity (e.g., as in typical alignment software such as BLAST) and then selects a set of overlapping oligonucleotides that encompass the regions of similarity and diversity, optionally using any 20 of the weighting factors described herein (e.g., predominant selection of oligonucleotides corresponding to one or more nucleic acid to be recombined, as in the gene blending methods herein). The computer or computer readable medium optionally comprises features 25 facilitating use by a user; e.g., an input field for inputting oligonucleotide selections by the user, a display output system for controlling a user-viewable output (e.g., a GUI), an output file which directs synthesis of the oligonucleotides, e.g., in an automated synthesizer, and the like.

EXAMPLE: BETALACTAMASE SHUFFLING WITH THREE BRIDGING OLIGOS

In this example, two beta lactamase genes (CFAMPC and FOX) were shuffled using three bridging oligonucleotides. The oligos were as follows:

- 30 1) CAAATACTGGCCGGAAC TGAAAGGTTCTGCTTCGACGGT
- 2) GTCGTGTTCTGCAGCCGCTGGGTCTGCACCACACCTACAT
- 3) TCGTTACTGGCGTATCGGTGACATGACCCAGGGTCTGGGT

The recombination reaction was performed using 2 micrograms of DNAsed 35 fragments from CFAMPC and FOX. All three oligos were added to the reaction 1:1 in a total of 60 microliters of 1x Taq-mix (7070 microliters of H₂O, 100 microliters Taq buffer, 600 microliters MgCl₂ (25 mM), 80 microliters dNTPs (100mM)).

Familly shuffling is done using the assembly conditions described in Crameri et al. (1995) *Nature* 391: 288-291, except that oligos are spiked into the assembly mix as described in Crameri et al. (1998) *Bio techniques* 18(2): 194-196. The PCR reactions with outside primer 1 for ATGATATTGATGATTA and 1 rev

ACAGATGCGTA
CRV2-12 TAAATCTATGACACTCATGAAATGGTACATTGATGATTCACTA
CRV2-13 TATACAGGATTACCTACATGCTTACATGCGACACTCAATTGCGA
CRV2-14 CAAATTGAGCTTACATTGCTTACATGCGAAATAGGATATTCTCGA
CRV2-15 AGATTCAGCTAAAGCCTAACCGGCTGATCAGCTTACACCTTAA
TGGAGATTTGAA

diversity of this region. Oligos are resuspended in 200 μ l H_2O . The oligos are as follows:

Genes to be used are *cry 2Aa*, *cry2Ab*, and *cry2Ac*. DNA alignment was done with DNA star using editseq, and megalign. Oligos are 50 nmoL synthesis (BRL, LifeTech.) with region between amino acid 260-630 are designed for *cry2Ac* in regard to

EXAMPLE: CREATING SEMISINTHETIC LIBRARY BY OLIGO SPlicing

digested with *Sph*I overnight at 30°C, gel purified and ligated into vector *Sh-BLA-Sh* (MG18), transformed into *TG1* and plated on *Tet* 20. 50 colonies were selected from the *Tet* 20 plates and amplified by colony PCR. The PCR amplicon was then digested overnight at 37°C with *Hinf*I. Restriction analysis revealed that 2 wt sequences for each parental gene, as well as 7 different recombinant products (for the 10 X molar reaction) or 8 different clones (for the 20 reaction) were produced.

(10 X molar), and 1500 ng primers (20X molar). 20 microliters of the assembling mix were added to 60 microliters of the λ Tag mix and 40 thermal cycles were performed at 94 °C (30 sec.), 40 °C (30 sec) and 72 °C (30 sec). 1, 2, 4, and 8 microliters of the resulting products were then PCR amplified for 40 cycles (same thermal cycling conditions as before) using primers for the end regions of the betalactamase genes. The resulting material was then

TTAATAAAAGTGGTGGAAAGATT are done with Taq /Pfu (9:1) mix (Taq from Qiagen, Pfu from Stratagene) PCR program 96°C (30 sec). 50°C (30sec). 72°C (1 min) for 25 cycles. The reaction is diluted 10x and an additional cycle is performed. The gene is ligated into a vector and transformed into TG1 competent Cells, and plated on LB +Amp100 plates. Single 5 colonies are picked for colony PCR and then analyzed by restriction digestion.

EXAMPLE: OLIGO SHUFFLING OF LIBRARIES

An advantage of oligonucleotide mediated shuffling methods is the ability to recombine nucleic acids between libraries of oligos generated for a number of different sites in a gene of interest. Generating libraries with complex combinations of randomizations in 10 different regions of a target gene is facilitated by oligonucleotide mediated shuffling approaches.

For example, the antigen-binding site of an antibody or antibody fragment such as a single-chain Fv (ScFv), or Fab is mainly comprised of 6 complementarity-determining regions (CDR's). These CDR's are present on one face of the properly folded 15 molecule, but are separated in the linear gene sequence. Synthetic oligonucleotides or those generated by PCR of one or more antibody genes can be used to generate sequence diversity at individual CDR's. This process can be repeated with a second CDR, then a third, until a library of diverse antibodies is formed. DNA shuffling formats have a distinct advantage that allow for libraries of each CDR to be generated simultaneously and inter-CDR recombination 20 events will frequently occur to potentially generate all possible combinations of different CDR's. Recursive DNA shuffling and screening for an improved trait or property can be used to optimize protein function.

Similarly, the 3-dimensional structures of many cytokines share a common 4-helix bundle structure with long connecting loops. The receptor binding sites for some of 25 these proteins has been determined and is localized to 2 or more regions of the protein that are separate in the linear gene sequence. Modeling of related proteins could be used to predict functional regions of unknown proteins for targeting libraries. Libraries in each of these regions can be generated using synthetic oligos, family-shuffling oligos, fragments of homologous genes, or combinations thereof as herein. Oligonucleotide mediated shuffling 30 allows one to generate libraries in each of these regions simultaneously and to generate recombinants between each library. In this way, combinations between members of each library can be screened for improved function. Those isolates with improved function can then be submitted to successive rounds of DNA shuffling. In this way, isolates with the

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the disclosure or kit herein.

30 Any apparatus or kit to practice any assay or method herein, for the practice of any method or assay herein, and/or for the use of any component or kit herein, for the practice of any method or assay herein, and/or for the use of any apparatus or kit herein, the present invention provides for the use of any

25 In a further aspect, the present invention provides for the use of any upon hybridization and elongation, will result in shuffled forms of the target nucleic acids, instruction sets for aligning target nucleic acids and for selecting oligonucleotides which, like (5) packaging materials, and (6) a computer or comparable medium having

holding nucleic acids or enzymes, other nucleic acids, transgenic plants, animals, cells, or the gene selection procedures herein; (3) one or more assay components; (4) a container for

the methods described herein, and/or for operating the oligonucleotide synthesis or assembled following: (1) a recombination component as described herein; (2) instructions for practicing methods and apparatus herein. Kits of the invention optionally comprise one or more of the

20 In an additional aspect, the present invention provides kits embodying the one or more containers which contain assay, device or system components, or the like, additionally comprising instructions for performing methods or assays, packaging materials, materials, components, methods or substrates hereinbefore described. Kits will optionally an assay, kit or system utilizing a use of any one of the selection strategies,

15 The use of an integrated system to select family shuffled oligonucleotides (e.g., by a process which includes sequence alignment of parental nucleic acids) and to test shuffled nucleic acids for activity, including in an iterative process.

The use of an integrated system to select family shuffled oligonucleotides invention can be put to a number of different uses, including:

10 Modifications can be made to the method and materials as hereinbefore described without departing from the spirit or scope of the invention as claimed, and the can be evaluated for an improved trait or property.

5 Substrate binding site(s) is comprised of residues close together in the 3-dimensional structure of the folded protein, but separated in the linear sequence of the gene. DNA shuffling can simultaneously generate libraries in each region that interact with substrate. DNA shuffling also allows all possible combinations of changes between each library to be generated and

Another example is the shuffling of enzymes where the active site and highest activity in each library and potential synergies between members of different libraries can be selected. Other methods that optimize each library independently may fail to isolate such synergistic interactions.

WO 00/42561 PCT/US00/01203

true scope of the invention. For example, all the techniques and materials described above can be used in various combinations. *All publications and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted.*

1. A method of recombining homogeneous nucleic acids, the method comprising:

(i) hybridizing a set of family gene shuffling oligonucleotides, thereby providing a population of recombinant nucleic acids;

(ii) elongating the set of family gene shuffling oligonucleotides;

(iii) elongating the set of family gene shuffling oligonucleotides, thereby providing a population of recombinant nucleic acids;

(iv) polymerase or a ligase;

3. The method of claim 1, wherein the elongating step is performed with a polymerase or a ligase.

4. The method of claim 1, wherein the set of family gene shuffling oligonucleotides encodes an evolutionary intermediate nucleic acid.

5. The method of claim 1, the method further comprising:

(i) recombining the denatured recombinant nucleic acids;

(ii) denaturing the population of recombinant nucleic acids, thereby providing a population of recombinant nucleic acids;

(iii) denaturing the set of family gene shuffling oligonucleotides;

4. The method of claim 1, wherein the set of family gene shuffling oligonucleotides are overlapping.

2. The method of claim 1, wherein the set of family gene shuffling oligonucleotides are overlapping.

5. (i) hybridizing a set of family gene shuffling oligonucleotides, thereby providing a population of recombinant nucleic acids;

(ii) elongating the set of family gene shuffling oligonucleotides;

3. The method of claim 1, wherein the elongating step is performed with a polymerase or a ligase.

4. The method of claim 1, wherein the set of family gene shuffling oligonucleotides encodes an evolutionary intermediate nucleic acid.

5. The method of claim 1, the method further comprising:

(i) recombining one or more of the resulting recombinant nucleic acids for a desired property;

(ii) recombining the resulting recombinant nucleic acids;

(iii) denaturing the resulting recombinant nucleic acids, thereby providing a population of recombinant nucleic acids;

(iv) recombining the resulting recombinant nucleic acids for a desired trait or property;

(v) extending or ligating the resulting recombinant nucleic acids; and,

(vi) selecting one or more of the resulting recombinant nucleic acids for a desired property.

6. The method of claim 5, wherein, prior to performing step (vi), the resulting recombinant nucleic acids are recombinant.

7. The method of claim 5, further comprising:

(vii) recombining the resulting recombinant nucleic acids;

8. The method of claim 7, further comprising selecting the resulting multiply selected multiply recombinant nucleic acids for a desired trait or property;

9. The method of claim 1, the method further comprising the steps of:

(iii) denaturing the population of recombinant nucleic acids, thereby providing a population of recombinant nucleic acids;

(iv) recombining the denatured nucleic acids;

30 (iv) recombining the denatured nucleic acids;

(v) extending the resulting reannealed recombined nucleic acids; and, repeating steps iii-v at least once.

10. The method of claim 1, further comprising selecting one or more of the resulting reannealed recombined nucleic acids for a desired trait or property.

5 11. The method of claim 1, further comprising selecting one or more member of the population of recombined nucleic acids for a desired property.

12. The method of claim 11, wherein a plurality of members of the population of recombined nucleic acids are screened for a desired property and are determined to have the desired property, thereby providing first round screened nucleic acids,

10 the method further comprising:

hybridizing a second set of family gene shuffling oligonucleotides, which second set of family gene shuffling oligonucleotides are derived from the first round screened nucleic acids; and,

15 elongating the second set of family gene shuffling oligonucleotides, thereby providing a population of further recombined nucleic acids.

13. The method of claim 12, wherein the second set of family gene shuffling oligonucleotides are overlapping.

14. The method of claim 12, further comprising sequencing the first round screened nucleic acids, wherein the second set of family gene shuffling oligonucleotides is derived from the first round screened nucleic acids by aligning sequences of the first round screened nucleic acids to identify regions of identity and regions of diversity in the first round screened nucleic acids, and synthesizing the second set of family gene shuffling oligonucleotides to comprise a plurality of oligonucleotides, each of which comprise subsequences corresponding to at least one region of diversity.

25 15. The method of claim 12, wherein the first round screened nucleic acids encode polypeptides of about 50 amino acids or less.

16. The method of claim 12, wherein the second set of family shuffling gene oligonucleotides comprise a plurality of oligonucleotide member types which comprise consensus region subsequences derived from a plurality of the first round screened nucleic acids.

acid recombination, the method comprising:

27. A method for introducing nucleic acid family diversity during nucleic acid homologous oligonucleotide member types are present in non-equimolar amounts, the homologous oligonucleotide member types are present in non-equimolar amounts, wherein the oligonucleotides comprises a plurality of homologous oligonucleotide member types, wherein the oligonucleotide member types are present in approximately equimolar amounts, the method of claim 1, wherein the set of family shuffling gene

26. The method of claim 1, wherein the set of family shuffling gene

oligonucleotide member types are present in approximately equimolar amounts, the oligonucleotide member types are present in approximately equimolar amounts, wherein the oligonucleotides comprising a plurality of homologous oligonucleotide member types, wherein the oligonucleotide member types are present in approximately equimolar amounts, the method of claim 1, the set of family shuffling gene oligonucleotides

25. The method of claim 1, the set of family shuffling gene oligonucleotides comprising a plurality of oligonucleotide member types comprising at least 10 member types, the method of claim 1, the set of family shuffling gene oligonucleotides

24. The method of claim 1, the set of family shuffling gene oligonucleotides comprising a plurality of oligonucleotide member types comprising at least 5 member types, the method of claim 1, the set of family shuffling gene oligonucleotides

23. The method of claim 1, the set of family shuffling gene oligonucleotides comprising a plurality of oligonucleotide member types comprising at least 3 member types, the method of claim 1, the set of family shuffling gene oligonucleotides

22. The method of claim 1, the set of family shuffling gene oligonucleotides comprising a plurality of codon-varied oligonucleotides, the method of claim 1, wherein the set of family shuffling

oligonucleotides comprise a plurality of codon-varied oligonucleotides, the method of claim 1, wherein the set of family shuffling

21. The method of claim 1, wherein the set of family shuffling

oligonucleotide module and a second sequence from a second sequence module, the method of claim 1, wherein the set of family shuffling

oligonucleotide module and a second sequence from a second sequence module, the method of claim 1, wherein the set of family shuffling

oligonucleotides comprising a plurality of module shuffling oligonucleotides, each comprising one or more of the plurality of oligonucleotides each comprise at least a first sequence from a first sequence module and a second sequence from a second sequence module, the method of claim 1, wherein the set of family shuffling

oligonucleotides comprising a plurality of module shuffling oligonucleotides, each comprising one or more of the plurality of oligonucleotides each comprise at least a first sequence from a first sequence module and a second sequence from a second sequence module, the method of claim 1, wherein the set of family shuffling

oligonucleotides comprising a plurality of module shuffling oligonucleotides, each comprising one or more of the plurality of oligonucleotides each comprise at least one module shuffling oligonucleotide(s), the method of claim 1, wherein the set of family shuffling gene

oligonucleotides comprising at least one module shuffling oligonucleotide(s), the method of claim 1, wherein the set of family shuffling gene

consensus region subsequences derived from a plurality of homologous target nucleic acids, the method of claim 1, wherein the set of family shuffling gene

oligonucleotides comprising a plurality of oligonucleotide member types which comprise

17. The method of claim 1, wherein the set of family shuffling gene

providing a composition comprising at least one set of fragmented nucleic acids and a population of family gene shuffling oligonucleotides;

recombinant at least one of the family gene shuffling oligonucleotides with at least one of the fragmented nucleic acids of the at least one set of fragmented nucleic acids; and,

5 regenerating a recombinant nucleic acid, thereby providing a regenerated recombinant nucleic acid comprising a nucleic acid subsequence corresponding to the at least one family gene shuffling oligonucleotide.

28. The method of claim 27, wherein the recombinant nucleic acid is selected for one or more desired trait or property.

10 29. The method of claim 28, wherein a plurality of members of recombined nucleic acids are screened for a desired property and are determined to have the desirable property, thereby providing first round screened nucleic acids, the method further comprising:

15 hybridizing a second set of overlapping family gene shuffling oligonucleotides, which second set of overlapping family gene shuffling oligonucleotides are derived from the first round screened nucleic acids; and,

elongating the second set of overlapping family gene shuffling oligonucleotides, thereby providing a population of further recombined nucleic acids.

20 30. The method of claim 29, further comprising sequencing the first round screened nucleic acids, wherein the second set of overlapping family gene shuffling oligonucleotides is derived from the first round screened nucleic acids by aligning sequences of the first round screened nucleic acids to identify regions of identity and regions of diversity in the first round screened nucleic acids, and synthesizing the second set of overlapping family gene shuffling oligonucleotides to comprise a plurality of oligonucleotides, each of which comprise subsequences corresponding to at least one region of diversity.

25 31. The method of claim 29, wherein the second set of overlapping family gene shuffling oligonucleotides comprise a plurality of oligonucleotide member types which comprise consensus region subsequences derived from a plurality of the first round screened nucleic acids.

30 32. The method of claim 27, wherein the set of overlapping family gene shuffling oligonucleotides comprise at least one module shuffling oligonucleotide(s).

30. 40. The method of claim 36, wherein the method further comprising selecting at least a second full-length nucleic acid and cleaving it to provide a second set of nucleic acid fragments by partial chain elongation.

38. 39. The method of claim 36, wherein the full-length nucleic acid is fragmented by cleavage with a DNase enzyme.

25. 38. The method of claim 36, wherein the full-length nucleic acid is fragmented by cleavage with a DNase enzyme, mixing the resulting set of nucleic acid fragments with the plurality of family gene shuffling oligonucleotides, thereby providing the composition comprising a fragmenting nucleic acid and a population of family gene shuffling oligonucleotides.

20. 39. The method of claim 36, wherein the full-length nucleic acid is identical to, or homologous with, at least one region of sequence diversity, providing a full-length nucleic acid which is identical to, or homologous with, at least one region of sequence diversity, least one region of sequence diversity, synthesizing a plurality of family gene shuffling oligonucleotides corresponding to at least one region of sequence diversity, identifying and regions of sequence diversity, aligning homologous nucleic acid sequences to select conserved regions of sequence provided by the steps of:

15. 37. The method of claim 27, wherein the composition comprising at least one fragmented nucleic acid and a population of family gene shuffling oligonucleotides is acid encodes a full-length protein.

36. 36. The method of claim 27, wherein the regenerated recombinant nucleic oligonucleotides comprise a plurality of codon-varied oligonucleotides.

10. 35. The method of claim 27, wherein the set of overlapping family shuffling gene oligonucleotides comprise a plurality of module shuffling oligonucleotides, wherein one or more of the plurality of oligonucleotides each comprising at least a first sequence from a first sequence module and a second sequence from a second sequence module.

5. 34. The method of claim 27, wherein the set of overlapping family shuffling subsequence from a second sequence module comprising at least a first subsequence from a first sequence module and a second subsequence from a second sequence module.

33. 33. The method of claim 27, wherein the set of overlapping family shuffling gene oligonucleotides comprise a plurality of module shuffling oligonucleotides, each comprising at least a first subsequence from a first sequence module and a second subsequence from a second sequence module.

fragments, which second set of nucleic acid fragments is also mixed with the population of gene shuffling oligonucleotides.

41. The method of claim 27, wherein the family gene shuffling oligonucleotides are provided to the composition by:

5 aligning homologous nucleic acid sequences and selecting at least one conserved region of sequence identity and a plurality of regions of sequence diversity, wherein the plurality of regions of sequence diversity provide a plurality of domains of sequence diversity; and,

10 synthesizing a plurality of family gene shuffling oligonucleotides corresponding to the plurality of domains of sequence diversity.

42. The method of claim 40, wherein recombination of the plurality of family gene shuffling oligonucleotides corresponding to the plurality of domains of sequence diversity with the fragmented nucleic acid causes domain switching in the regenerated recombinant nucleic acid, as compared to the homologous nucleic acid sequences.

15 43. The method of claim 40, wherein the plurality of family gene shuffling oligonucleotides corresponding to the plurality of domains of sequence diversity is synthesized by synthesizing family gene shuffling oligonucleotides which encode one or more domain of sequence diversity corresponding to one or more of the homologous nucleic acid sequences.

20 44. The method of claim 27, wherein the fragmented nucleic acid is provided by one or more of: (i) cleaving a cloned nucleic acid, and (ii) selecting a nucleic acid sequence and synthesizing oligonucleotide fragments corresponding to the selected nucleic acid sequence.

25 45. A method of recombining homologous or non-homologous nucleic acid sequences having low sequence similarity, the method comprising:

recombining one or more set of fragmented nucleic acids with a set of crossover oligonucleotides, which oligonucleotides individually comprise a plurality of sequence diversity domains corresponding to a plurality of sequence diversity domains from homologous or non-homologous nucleic acids with low sequence similarity, thereby 30 producing a recombinant nucleic acid.

46. The method of claim 44, further comprising selecting the recombinant nucleic acid for a desired trait or property.

30 56. A method of family shuffling PCR amplicons, the method comprising:
 to one or more of the homologous nucleic acid sequences.
 member of the oligonucleotide set with one or more homologous nucleic acid corresponding
 member of the oligonucleotide set comprising recombining one or more
 25 55. The method of claim 49, further comprising combining one or more
 oligonucleotides, resulting from recombining the oligonucleotide set, for a desired trait or
 54. The method of claim 52, further comprising selecting any recombinant
 oligonucleotide set.
 20 53. The method of claim 49, further comprising recombining the
 wherein the homologous sequences are aligned by manual alignment.
 are aligned in a system comprising a computer with software for sequence alignment, or
 52. The method of claim 49, wherein the homologous nucleic acid sequences
 member types are synthesized serially or in parallel.
 15 51. The method of claim 49, wherein the plurality of oligonucleotide
 sequence heterogeneity corresponding to at least one of the homologous nucleic acids.
 oligonucleotides which comprise at least one member type comprising at least one region of
 to at least one of the one or more regions of heterogeneity, thereby providing a set of
 synthesizing a plurality of different oligonucleotide member types which correspond
 region of sequence heterogeneity; and,
 10 aligning a plurality of homologous nucleic acid sequences to identify one or more
 homologous nucleic acids, the method comprising:
 50. A method of providing an oligonucleotide set for recombination of
 homologous nucleic acid, thereby providing the one or more fragmented nucleic acid.
 49. The method of claim 44, the method further comprising synthesizing a
 homologous nucleic acid is fragmented with a DNase enzyme.
 48. The method of claim 46, wherein the one or more homologous or non-
 nucleic acids.
 47. The method of claim 44, the method further comprising fragmenting one
 or more of the homologous or non-homologous nucleic acids to provide the set of fragmented
 5 nucleic acids.

providing a plurality of non-homogeneous homologous template nucleic acids;

providing a plurality of PCR primers, which PCR primers hybridize to a plurality of the plurality of non-homogeneous homologous template nucleic acids;

5 producing a plurality of PCR amplicons by PCR amplification of the plurality of template nucleic acids with the plurality of PCR primers; and,

recombinant the plurality of PCR amplicons, thereby providing a recombinant nucleic acid.

57. The method of claim 55, further comprising selecting the recombinant nucleic acid.

10 58. The method of claim 55, wherein a sequence for the PCR primers is selected by aligning sequences for the plurality of non-homogeneous homologous template nucleic acids, and selecting PCR primers which correspond to regions of sequence similarity.

59. A method of recombinant a plurality of parental nucleic acids, the method comprising:

15 ligating or elongating a set of a plurality of oligonucleotides, the set comprising a plurality of nucleic acid sequences from a plurality of the parental nucleic acids to produce a recombinant nucleic acid encoding a full length protein.

20 60. The method of claim 59, the set comprising at least a first oligonucleotide which is complementary to at least a first of the parental nucleic acids at a first region of sequence diversity and at least a second oligonucleotide which is complementary to at least a second of the parental nucleic acids at a second region of diversity.

61. The method of claim 59, wherein the nucleic acids are ligated with a ligase.

25 62. The method of claim 59, wherein the oligonucleotides are hybridized to a first parental nucleic acid and ligated with a ligase.

63. The method of claim 59, wherein the parental nucleic acids are homologous.

30 64. The method of claim 59, wherein the set of oligonucleotides comprises a set of family gene shuffling oligonucleotides.

65. The method of claim 59, the method further comprising hybridizing the set of oligonucleotides to one or more of the parental nucleic acids, and elongating the oligonucleotides with a polymerase to produce a nucleic acid encoding a substantially full-length protein.

66. A method of producing a recombinant nucleic acid, the method comprising:

(i) transducing a population of cells with a set of overlapping family gene shuffling oligonucleotides;

(ii) permitting recombination to occur between the set of overlapping family gene oligonucleotides; and,

67. The method of claim 66, further comprising selecting the population of recombinant cells for a desired trait or property.

68. The method of claim 66, further comprising PCR amplifying the population of recombinant nucleic acids.

69. The method of claim 68, further comprising transducing the PCR amplifying the population of recombinant nucleic acids into a cell, vector, or virus.

70. The method of claim 66, wherein the set of overlapping family gene shuffling oligonucleotides are chimeraplasts.

71. The method of claim 70, wherein the chimeraplasts are codon-varied

72. The method of claim 64, wherein the set of overlapping family gene shuffling oligonucleotides comprises a plurality of codon-varied oligonucleotides.

73. The population of recombinant nucleic acids produced by the method of claim 66.

74. The population of recombinant cells produced by the method of claim 66.

75. An amplified nucleic acid produced by the method of claim 68.

76. A cell, vector, or virus produced by the method of claim 69.

77. A composition comprising a library of oligonucleotides comprising a plurality of oligonucleotide member types, the oligonucleotide member types corresponding to a plurality of subsequence regions of a plurality of members of a selected set of a plurality of homologous target sequences.

5 78. The composition of claim 77, wherein the library comprises at least about 10, 20, 30, 40, 50 or more different oligonucleotide members.

79. The composition of claim 77, wherein the oligonucleotide member types are present in non-equimolar amounts.

10 80. The composition of claim 77, the plurality of subsequence regions comprising a plurality of non-overlapping sequence regions of the selected set of homologous target sequences.

81. The composition of claim 77, wherein the oligonucleotide member types each have a sequence identical to at least one subsequence from at least one of the selected set of homologous target sequences.

15 82. The composition of claim 77, wherein the oligonucleotide member types comprise a plurality of homologous oligonucleotides corresponding to a homologous region from the plurality of homologous target sequences, wherein each of the plurality of homologous oligonucleotides comprise at least one variant subsequence.

20 83. The composition of claim 77, further comprising one or more of: a polymerase, a thermostable DNA polymerase, a nucleic acid synthesis reagent, a buffer, a salt, magnesium, and one or more nucleic acid comprising one or more of the plurality of members of the selected set of homologous target sequences.

25 84. The composition of claim 77, wherein the plurality of oligonucleotide member types is selected by aligning the plurality of homologous target sequences, determining at least one region of identity and at least one region of variance and synthesizing the oligonucleotides to encode at least a portion of the at least one region of identity, or at least a portion of the at least one region of variance, or at least a portion of both the at least one region of identity and at least one region of variance.

30 85. The composition of claim 77, wherein the plurality of oligonucleotide member types comprise at least one member type comprising at least one sequence diversity domain.

30. 95. The method of claim 94, further comprising selecting the resulting further recombinant nucleic acid for a desired trait or property.

30. 94. The method of claim 89, further comprising recombinant nucleic acid with an additional nucleic acid.

30. 93. The method of claim 92, further comprising repeating any of steps (i)-(iv).

25. 92. The method of claim 89, further comprising: (i.) selecting the plurality of recombinant nucleic acids for a desired trait or property.

25. 91. The method of claim 89, wherein the two or more nucleic acids are non-homologous.

20. 90. The method of claim 89, wherein the two or more nucleic acids are of recombinant nucleic acids.

15. 89. A method of recombinant nucleic acids to identify regions of identity, (i.) aligning two or more nucleic acids to identify regions of identity and regions of diversity; (ii.) providing a non-equimolar set of oligonucleotides which comprise a plurality of oligonucleotides which correspond in sequence to at least two of the two or more nucleic acids at least one region of diversity, the oligonucleotides being present in non-equimolar amounts; and, (iii.) extending the oligonucleotides with a polymerase, thereby producing a plurality of recombinant nucleic acids.

10. 88. The composition of claim 86, wherein the sequence diversity domains correspond to adjacent sequence regions on a plurality of the homologous nucleic acids when the homologous nucleic acids are aligned.

5. 87. The composition of claim 77, wherein the library comprises a set of crossover family diversity oligonucleotides, each oligonucleotide member of the set of crossover family diversity oligonucleotides comprising a plurality of sequence diversity domains corresponding to a plurality of homologous nucleic acids.

86. The composition of claim 77, wherein the plurality of oligonucleotide member types comprise a plurality of sequence diversity domains.

86. The composition of claim 77, wherein the plurality of oligonucleotide

96. A method of making a library of chimeraplasts, the method comprising:
providing a plurality of homologous chimeraplasts, each comprising a marker or other
region of sequence similarity, and at least one region of sequence difference, thereby
producing a library of chimeraplasts.

5 97. The method of claim 96, wherein the plurality of chimeraplasts are
codon-varied oligonucleotides.

98. The library produced by the method of claim 96.

99. The method of claim 96, further comprising transducing a population of
cells with the library of chimeraplasts and detecting recombination of the marker or other
10 region of similarity with one or more nucleic acid in the cell, and identifying which of the
homologous chimeraplasts recombined with the one or more nucleic acid in the cell, thereby
identifying active homologous chimeraplasts.

15 100. The method of claim 99, further comprising recombining a plurality of
the active homologous chimeraplasts to produce a library of recombined active homologous
chimeraplasts.

101. The library produced by the method of claim 100.

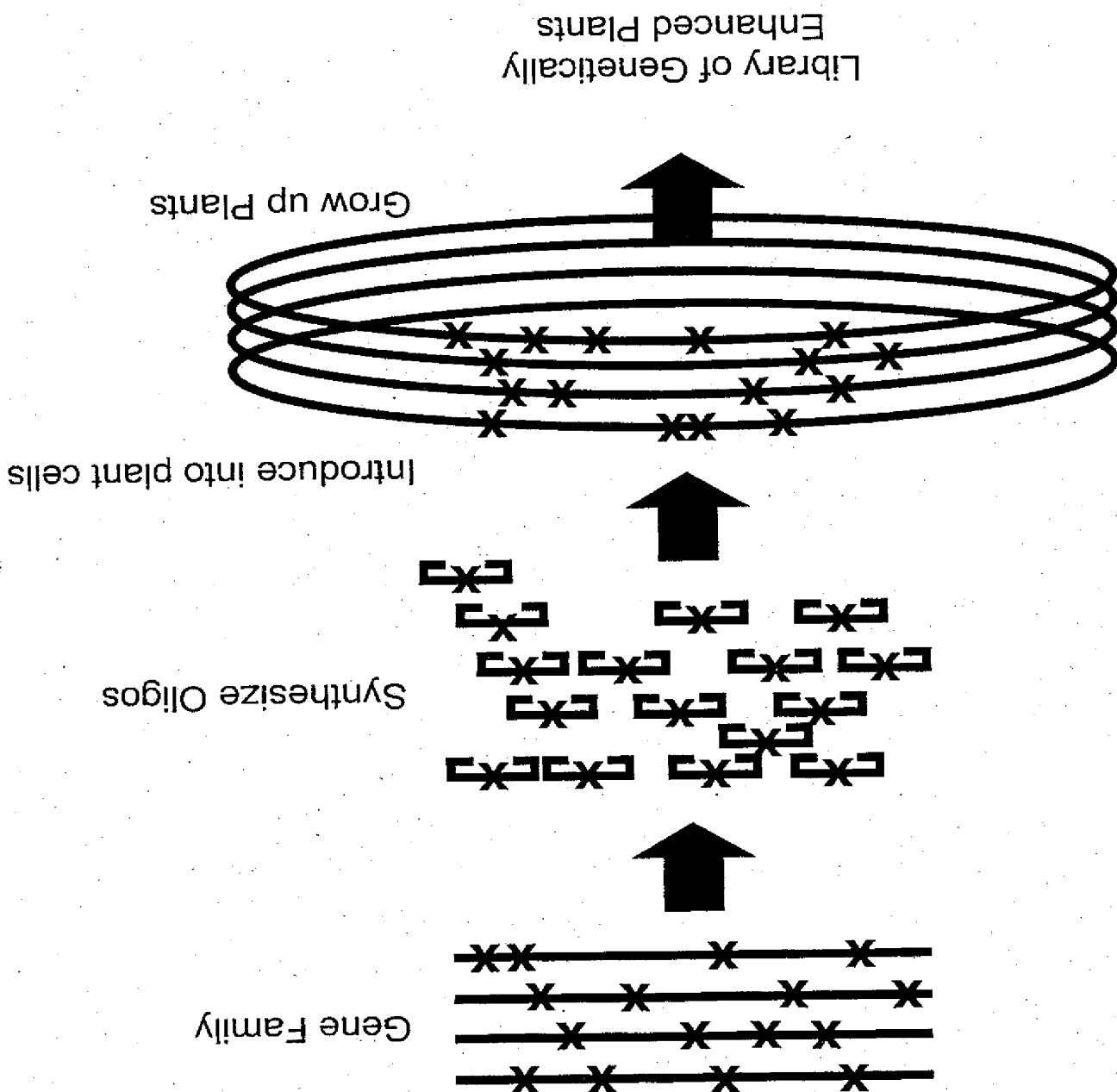
102. The method of claim 100 further comprising transducing a second
population of cells with the library of recombined active homologous chimeraplasts and
identifying which of the active homologous chimeraplasts recombined with the one or more
20 nucleic acid in the cell, thereby identifying additional active homologous chimeraplasts.

103. The method of claim 102, further comprising providing a library of the
additional active homologous chimeraplasts.

104. The library produced by the method of claim 103.

SUBSTITUTE SHEET (RULE 26)

Fig. 1



2/3

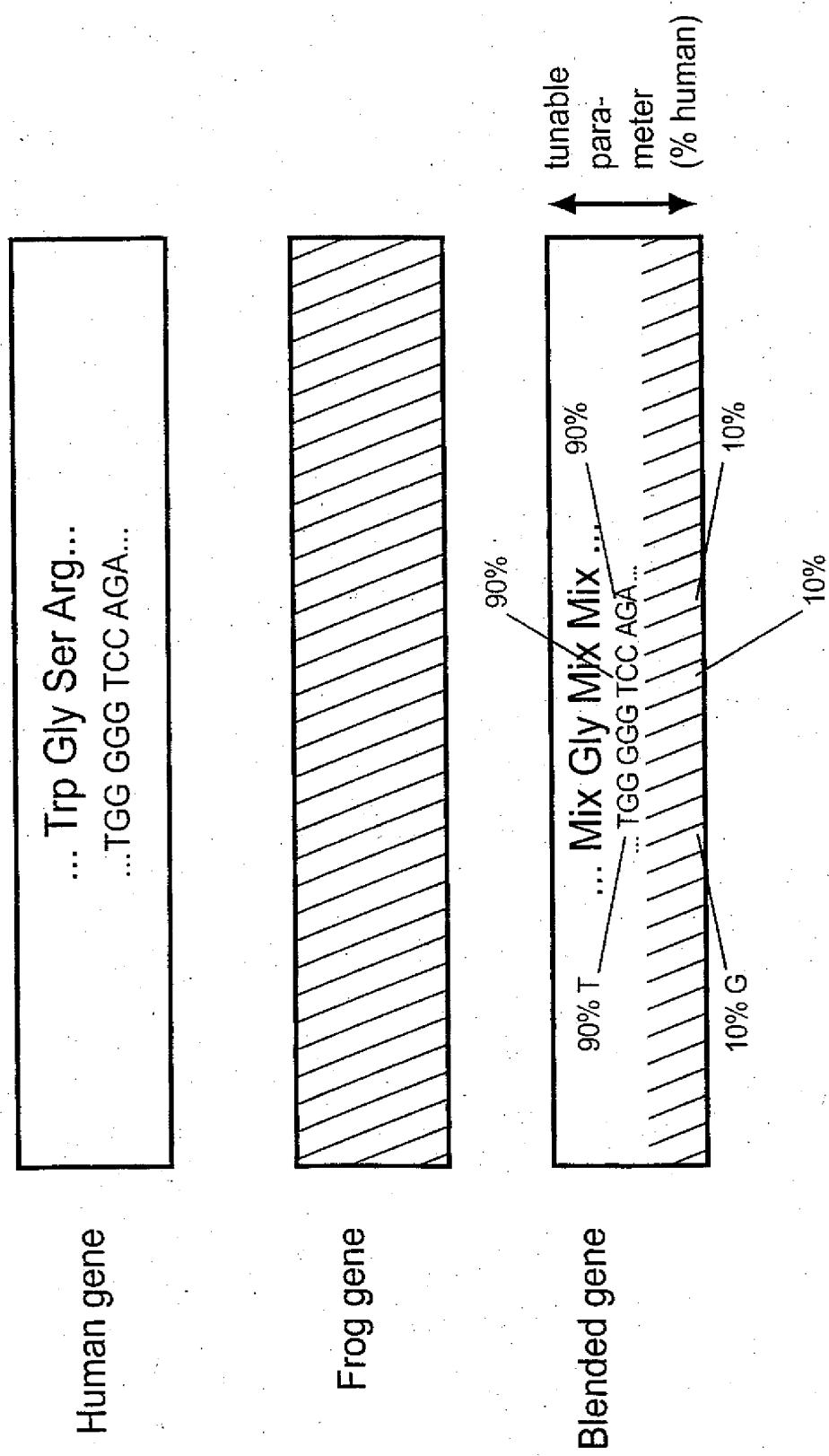
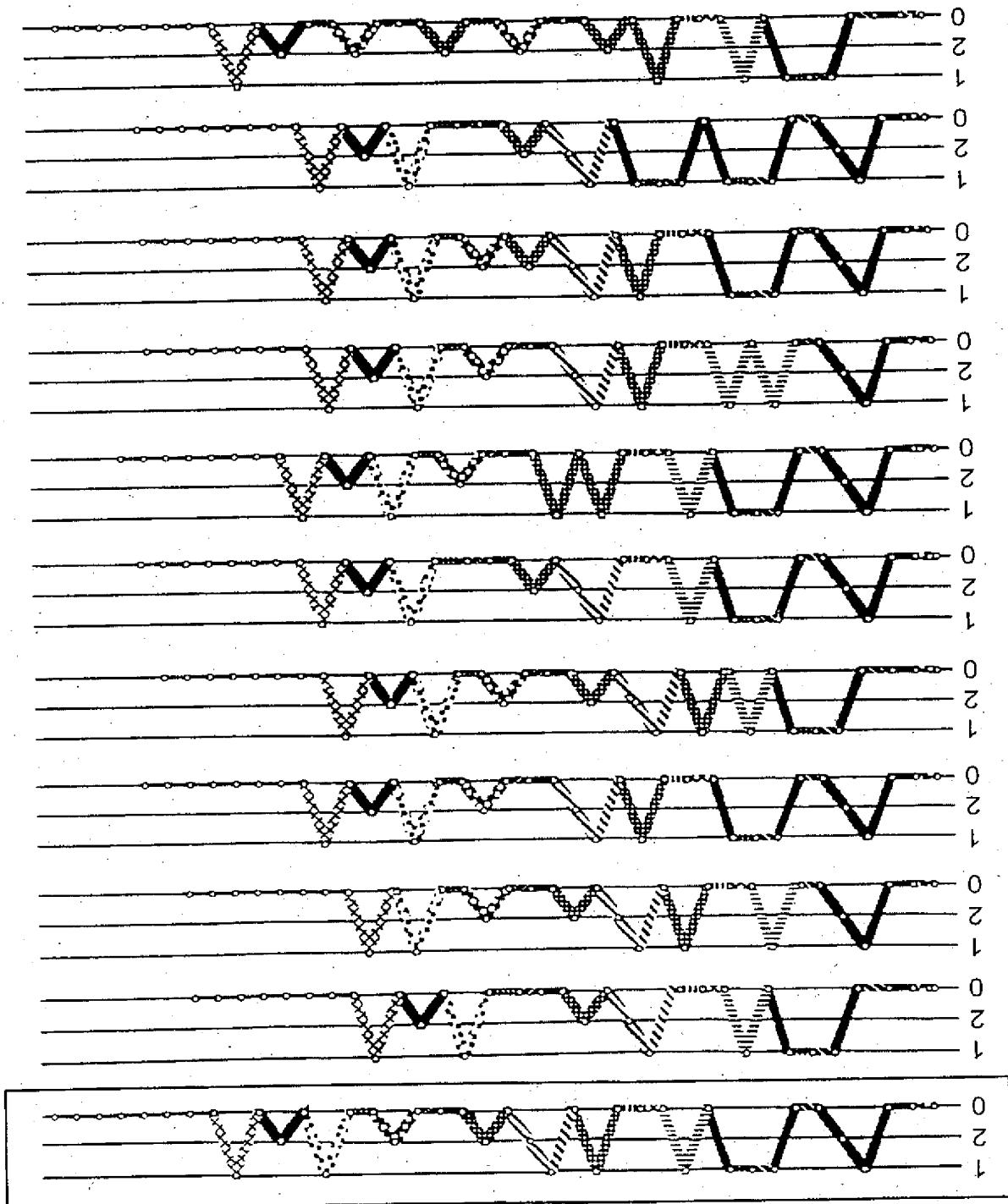


Fig. 2

SUBSTITUTE SHEET (RULE 26)

Fig. 3



3/3

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
20 July 2000 (20.07.2000)

PCT

(10) International Publication Number
WO 00/42561 A3

(51) International Patent Classification⁷: C12N 15/10, C12Q 1/68; G06F 19/00

(21) International Application Number: PCT/US00/01203

(22) International Filing Date: 18 January 2000 (18.01.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/116,447	19 January 1999 (19.01.1999)	US
60/118,813	5 February 1999 (05.02.1999)	US
60/118,854	5 February 1999 (05.02.1999)	US
60/141,049	24 June 1999 (24.06.1999)	US
09/408,392	28 September 1999 (28.09.1999)	US
09/408,393	28 September 1999 (28.09.1999)	US
09/416,375	12 October 1999 (12.10.1999)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/408,392 (CIP)
Filed on 28 September 1999 (28.09.1999)

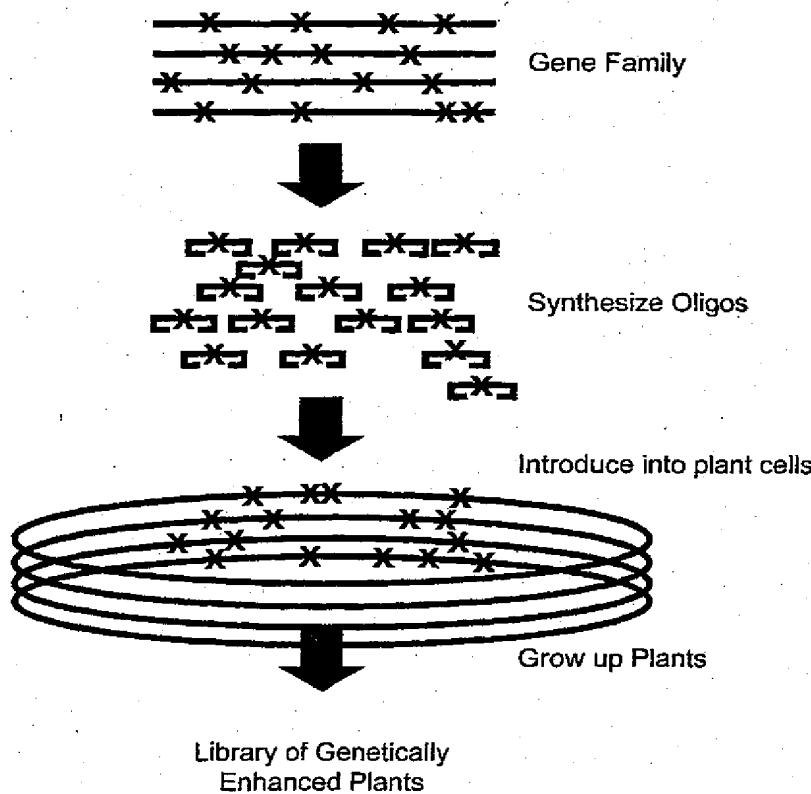
(71) Applicant (for all designated States except US): MAXY-GEN, INC. [US/US]; 515 Galveston Drive, Redwood City, CA 94063 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): CRAMERI, Andreas [CH/US]; Gehrstrasse 3, CH-4153 Reinach (CH). STEMMER, Willem, P., C. [NL/US]; 108 Kathy Court, Los Gatos, CA 95030 (US). MINSHULL, Jeremy [GB/US]; 11 Homer Lane, Menlo Park, CA 94025 (US). BASS, Steven, H. [US/US]; 950 Parrot Drive, Hillsborough, CA 94010 (US). WELCH, Mark [US/US]; 25 Montalban Drive, Fremont, CA 94536 (US). NESS, Jon,

[Continued on next page]

(54) Title: OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION



(57) Abstract: Methods of recombining nucleic acids, including homologous nucleic acids, are provided. Families of gene shuffling oligonucleotides and their use in recombination procedures, as well as polymerase and ligase mediated recombination methods are also provided.

WO 00/42561 A3

(84) Designated States (regional): ARIPO patent (GH, GM, EA, LS, MW, SD, SI, SZ, TZ, UG, ZW), Burasian patent (AM, AZ, BY, KG, KZ, MD, RU, TI, TM), European Patent (AT, BE, CH, CY, DE, DK, ES, FR, FR, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BI, CR, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(85) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TI, TM, TR, TT, TZ, UA, Ukraine, VN, YU, ZA, ZW.

(86) Date of publication of the international search report: 7 December 2000

(87) With international search report

(88) Date of publication of the international search report: 7 December 2000

(89) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TI, TM, TR, TT, TZ, UA, Ukraine, VN, YU, ZA, ZW.

(90) Two-letter codes and other abbreviations, refer to the "Guidelines of each regular issue of the PCT Gazette".



WO 00/42561 A3

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/01203

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 7 C12N15/10 C12Q1/68 G06F19/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 IPC 7 C12N C12Q G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 27230 A (MAXYGEN INC ;PATTEN PHILLIP A (US); STEMMER WILLEM P C (US)) 25 June 1998 (1998-06-25) claims 1,16,65	1-95
Y	page 14, paragraph 3 – paragraph 4 page 22, paragraph 1 – paragraph 3 page 28, line 14 –page 29, line 27 ---	96-104
Y	WO 95 15972 A (UNIV JEFFERSON) 15 June 1995 (1995-06-15) claims 12,15 page 8, line 28 –page 9, line 7 page 11, line 6 – line 15 ---	96-104
		-/-

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

30 August 2000

06/09/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patenttaan 2
NL – 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl
Fax: (+31-70) 340-3016

Authorized officer

Filloy Garcia, E

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		Category	Relevant to claim No.
1	International Application No. PCT/US 00/01203		
X	WO 98 42832 A (SHAO ZHIXIN ; ZHAO HUIMIN (CN); AFHOLTER JOSEPH A (US); ARNOLD FRA) 1 October 1998 (1998-10-01)	1. Category of document, with indication, where appropriate, of the relevant passages	
X	ZHAO H ET AL: "MOLECULAR EVOLUTION BY STAGGERED EXTENSION PROCESS (STEP) IN VITRO RECOMBINATION". NATURE BIOTECHNOLOGY, US, NATURE PUBLISHING, Vol. 16, 1 March 1998 (1998-03-01), pages 258-261, XP000775867 ISSN: 1087-0156	Abstract; Claims 1,17 Abstract 258, left-hand column, paragraph 3 page 259, right-hand column, paragraph 3	1,5,9,11
X	WO 97 20078 A (AFFYMAX TECH NV ; CRAMERI ANDREAS (US); STEMMER WILLEM P C (US)) 5 June 1997 (1997-06-05)	Abstract; Claims 1,17 Abstract 147, paragraph 1	1-104
A	GIVER L ET AL: "COMBINATORIAL PROTEIN DESIGN BY IN VITRO RECOMBINATION". CURRENT OPINION IN CHEMICAL BIOLOGY, 335-338, XP000892913 ISSN: 1367-5931	the whole document	96-104
	WO 98 49350 A (KREN BETSY T ; UNIV MINNESOTA (US); BANDYOPADHYAY PARAMITA T (US);) 5 November 1998 (1998-11-05)	Page 4, paragraph 3 Page 7, paragraph 5	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/01203

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9827230	A 25-06-1998	AU 5729298 A EP 0946755 A		15-07-1998 06-10-1999
WO 9515972	A 15-06-1995	AU 691550 B AU 1399595 A CA 2178729 A CN 1142829 A, B CN 1215755 A DE 733059 T EP 0733059 A JP 9506511 T NZ 278490 A US 5565350 A US 5756325 A US 5871984 A		21-05-1998 27-06-1995 15-06-1995 12-02-1997 05-05-1999 28-08-1997 25-09-1996 30-06-1997 25-03-1998 15-10-1996 26-05-1998 16-02-1999
WO 9842832	A 01-10-1998	AU 6772598 A AU 6942098 A BR 9804791 A CN 1230990 T EP 0975653 A EP 0920496 A PL 330287 A WO 9842728 A		20-10-1998 20-10-1998 17-08-1999 06-10-1999 02-02-2000 09-06-1999 10-05-1999 01-10-1998
WO 9720078	A 05-06-1997	US 5811238 A AU 713952 B AU 1087397 A AU 2542697 A CA 2239099 A EP 0876509 A EP 0906418 A EP 0911396 A JP 2000500981 T JP 2000507444 T WO 9735966 A US 5837458 A		22-09-1998 16-12-1999 19-06-1997 17-10-1997 05-06-1997 11-11-1998 07-04-1999 28-04-1999 02-02-2000 20-06-2000 02-10-1997 17-11-1998
WO 9849350	A 05-11-1998	AU 7365498 A EP 0979311 A		24-11-1998 16-02-2000

